No. 23-5600

UNITED STATES COURT OF APPEALS
FOR THE SIXTH CIRCUIT

L.W., by and through her parents and next friends Samantha Williams and Brian Williams, et al.,

Plaintiffs-Appellees,

v.

JONATHAN THOMAS SKRMETTI, in his official capacity as the Tennessee Attorney General and Reporter, et al.,

Defendants-Appellants,

and

UNITED STATES OF AMERICA,

Intervenor-Appellee.

On Appeal from the United States District Court
for the Middle District of Tennessee
Case No. 3:23-cv-00376

BRIEF OF ALLIANCE DEFENDING FREEDOM AS AMICUS CURIAE IN SUPPORT OF APPELLANTS AND FOR REVERSAL

JACOB P. WARNER
ALLIANCE DEFENDING FREEDOM
15100 N. 90th Street
Scottsdale, AZ 85260
(480) 444-0020
jwarner@ADFlegal.org

JOHN J. BURSCH
JAMES A. CAMPBELL
ALLIANCE DEFENDING FREEDOM
440 First Street NW
Suite 600
Washington, DC 20001
(616) 450-4235
jbursch@ADFlegal.org
jcampbell@ADFlegal.org

Attorneys for Amicus Curiae

UNITED STATES COURT OF APPEALS
FOR THE SIXTH CIRCUIT

Disclosure of Corporate Affiliations and Financial Interest

Sixth Circuit
Case Number: 23-5600 Case Name: L.W. v. Skrmetti

Name of counsel: John J. Bursch

Pursuant to 6th Cir. R. 26.1, Amicus, Alliance Defending Freedom
Name of Party
makes the following disclosure:

1. Is said party a subsidiary or affiliate of a publicly owned corporation? If Yes, list below the identity of the parent corporation or affiliate and the relationship between it and the named party:

No

2. Is there a publicly owned corporation, not a party to the appeal, that has a financial interest in the outcome? If yes, list the identity of such corporation and the nature of the financial interest:

No

CERTIFICATE OF SERVICE

I certify that on July 24, 2023 the foregoing document was served on all parties or their counsel of record through the CM/ECF system if they are registered users or, if they are not, by placing a true and correct copy in the United States mail, postage prepaid, to their address of record.

s/ John J. Bursch

This statement is filed twice: when the appeal is initially opened and later, in the principal briefs, immediately preceding the table of contents. See 6th Cir. R. 26.1 on page 2 of this form.

TABLE OF CONTENTS

TABLE OF AUTHORITIES

Cases

Statutes

Other Authorities

Regulations

IDENTITY AND INTEREST OF AMICUS CURIAE[1]

Alliance Defending Freedom is the world's largest law firm dedicated to religious freedom, free speech, the sanctity of life, parental rights, and marriage and family. Because the law should protect life—including from irreversible and unproven medical interventions—ADF advocates for laws that protect children from drug treatments that could potentially harm them for life.

ADF is deeply concerned about the use of puberty blockers and cross-sex hormones for children with gender dysphoria. Systematic reviews have shown insufficient evidence to support such use. Many studies even suggest this use is dangerous. This has led many European nations and American states to forbid puberty blockers and cross-sex hormones for children with gender dysphoria. ADF believes such caution is best, given the high stakes and uncertain science.

ADF has served as co-counsel defending states that protect children from potentially dangerous drug interventions, *e.g.*, *Boe v. Marshall*, No. 2:22-cv-184-LCB (M.D. Ala.), and submits this brief supporting Tennessee's Prohibition on Medical Procedures Performed on Minors Related to Sexual Identity, Tenn. Code Ann. 68-33-101 ("the Minors Protection Act" or simply the "Act").

---

[1]No counsel for a party authored this brief in whole or in part, and no person other than amicus and its counsel made any monetary contribution intended to fund the preparation or submission of this brief. All counsel were timely notified of this brief and consented to its filing.

INTRODUCTION

Tennessee seeks to protect children from unproven drug treatments that risk permanent harm. It enacted the Minors Protection Act to regulate puberty blockers and cross-sex hormones for children experiencing gender dysphoria. After examining the medical literature and best practices around the world, Tennessee found such drug use is harmful, as it can cause irreversible sterility, increase children's risk of disease and illness, and spark adverse and sometimes fatal psychological consequences. At minimum, the State found that using these drugs is reckless because they are experimental, unsupported by high-quality evidence, and pose unknown risks. Respondents challenge this protection, seeking a constitutional right to inject children with experimental drugs.

The district court preliminarily enjoined the Act, holding that it targets individuals who identify as transgender and lacks substantial basis. That ruling wrongly assumed that all individuals who suffer from gender dysphoria identify as transgender, and it credited expert testimony inconsistent with principles of evidence-based medicine—valuing low-quality anecdote over high-quality systematic reviews. The court should have found instead that no high-quality evidence supports puberty blockers and cross-sex hormones to treat children with gender dysphoria, and multiple studies suggest that such use is potentially dangerous.

Notably, courts give legislatures wide discretion to pass legislation when there is medical and scientific uncertainty. Just last year, the Supreme Court reversed a 50-year-old precedent constitutionalizing a right to abortion, recognizing it had improperly withheld judicial restraint on a critical social issue, causing great turmoil. This Court should not repeat that error here by constitutionalizing a new right to unproven medical treatments.

Accordingly, ADF asks this Court to reverse the ruling below and allow Tennessee to continue protecting its children.

ARGUMENT

An injunction "is an extraordinary remedy never awarded as of right." *Winter v. Nat. Res. Def. Council, Inc.*, 555 U.S. 7, 24 (2008). It "may only be awarded upon a *clear showing* that the plaintiff" deserves it. *Id.* at 22 (emphasis added). This is especially true when plaintiffs seek to enjoin the "enforcement of a presumptively valid state statute." *Brown v. Gilmore*, 122 S. Ct. 1, 1 (2001) (Rehnquist, C.J., in chambers). Such a request "demands" unusually strong "justification." *Lux v. Rodrigues*, 561 U.S. 1306, 1307 (2010) (Roberts, C.J., in chambers).

To obtain a preliminary injunction, Plaintiffs must prove, among other things, that they are "likely to succeed on the merits." *Winter*, 555 U.S. at 20. Plaintiffs have not done so here. This brief shows that Plaintiffs are unlikely to succeed because only rational-basis review applies to

their claims and the Act satisfies even intermediate scrutiny by reasonably protecting children from unproven drugs.

I. Rational-basis review applies, and the Act easily satisfies both rational-basis and intermediate scrutiny.

Statutory classifications are typically valid if they rationally advance a legitimate interest. *San Antonio Indep. Sch. Dist. v. Rodriguez*, 411 U.S. 1, 55 (1973). Closer scrutiny applies when laws implicate suspect or quasi-suspect classes. *Reed v. Reed*, 404 U.S. 71, 76 (1971). Laws that implicate sex or other quasi-suspect classifications must advance an "important" goal through "substantially related" means. *Tuan Anh Nguyen v. INS*, 533 U.S. 53, 60 (2001). But a perfect fit is not required; only a "reasonable" one. *Tyler v. Hillsdale Cnty. Sheriff's Dep't*, 837 F.3d 678, 693 (6th Cir. 2016). Rational-basis review applies here because the Minors Protection Act does not target a suspect or quasi-suspect class as it regulates drugs used on minors of both sexes. Regardless, the Act satisfies even intermediate scrutiny.

A. The Act protects children from unproven drug treatments no matter how they identify.

Tennessee enacted the Minors Protection Act "to protect the health and welfare of minors." Tenn. Code Ann. § 68-33-101(a). In its view, drugs "that alter a minor's hormonal balance" or otherwise change a "minor's physical" attributes "are harmful" when administered to children with gender dysphoria. *Id.* § 68-33-101(b). Such drugs "can lead to [a] minor

becoming irreversibly sterile," increase the "risk of disease and illness," and spark "adverse and sometimes fatal psychological consequences." *Id.* At minimum, Tennessee found that providing children these drugs is reckless because they "are experimental," "not supported by high-quality" evidence, and may have "harmful effects" not "fully known." *Id.*

The district court held that the Act "targets transgender people," because "inherent in a gender dysphoria diagnosis is a person's identity as transgender" and "a person cannot suffer from gender dysphoria without identifying as transgender. *L.W. by & through Williams v. Skrmetti*, No. 3:23-CV-00376, 2023 WL 4232308, at *10 (M.D. Tenn. June 28, 2023). Not so. Gender dysphoria is a recognized mental health condition. Am. Psychiatric Ass'n, *Diagnostic & Statistical Manual of Mental Disorders* 512 (5th ed. 2013). It requires six-month "marked incongruence between one's experienced/expressed gender and assigned gender" that is "associated with clinically significant distress." *Id.* In contrast, transgender *identification* is not a mental disorder. R.29, 250. People can identify as transgender without suffering from gender dysphoria. *Expert Q&A: Gender Dysphoria*, Am. Psychiatric Ass'n, https://perma.cc/3YJ4-F2A2 (last accessed July 13, 2023).

What's more, adolescent gender dysphoria often does not lead to adult transgender identification. Until recently, most minors presenting with gender dysphoria were pre-pubescent males. The Cass Review, *In-*

dependent review of gender identity services for children and young people: Interim report 32 (2022), https://perma.cc/9CT5-J6NU. The Dutch protocol analyzed this population. E. Abbruzzese et al., *The Myth of 'Reliable Research' in Pediatric Gender Medicine: A critical evaluation of the Dutch Studies—and research that has followed* 12, J. Sex & Marital Therapy (2023), App.94. With psychotherapy (but not social or medical transition practices), the study showed the vast majority of these children ceased to experience gender dysphoria during adolescence and identified with their natal sex as an adult. Wylie C. Hembree et al., *Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society Clinial Practice Guideline*, 102:11 J. Clinical Endocrinal Metab. 3869, 3879 (2017), App.726; James M. Cantor, *American Academy of Pediatrics policy and trans- kids: Fact-checking* 1, Sexology Today! (2018), App.464.

Such desistence is good. The affected individual no longer has a mental health condition and needs no further treatment. While the district court criticized Tennessee's aim to relieve children of "lifetime dependence" on experimental treatments, *Skrmetti*, 2023 WL 4232308 at *23 n.43, the Act seeks to prevent children from "iatrogenic" measures—treatments that *create* disease rather than cure it. Kenneth J. Zucker, *Debate: Different strokes for different folks* 1-2, Child & Adolescent Mental Health (2019), App.517-18. These drug interventions substantially

risk disrupting the ordinary resolution of gender dysphoria. In fact, children subject to such intervention are far more likely to *persist* in experiencing gender dysphoria than those who aren't. Kristina R. Olson et al., *Gender Identity 5 Years After Social Transition*, 150:2 Pediatrics 3 (2022), App.521. So early transition "is not a neutral act." Cass Review 38, 62-63. Tennessee's Act creates space for children's gender dysphoria to resolve.

This space is critical because a new population dominates gender clinics: females in mid-adolescence with gender discordance *without* a childhood history of such. Riittakerttu Kaltiala-Heino et al., *Two years of gender identity service for minors: overrepresentation of natal girls with severe problems in adolescent development*, 9 Child & Adolescent Psychiatry & Mental Health 6 (2015), App.615; Lisa Littman, *Parent reports of adolescents and young adults perceived to show signs of a rapid onset of gender dysphoria* 3, PLOS ONE (2018), https://perma.cc/E8ZH-FWP6; Cass Review 38. Early research did not study this population. Abbruzzese, *supra*, at 12, App.94. And modern research lags because this group is newly developing. Littman, *supra*, at 3; R.113-3, 1154-69. Nothing suggests this population will necessarily identify as transgender in adulthood. So caution is critical.

Tennessee enacted the Act to protect the health and safety of its children, no matter how they identify. The Act thus does not distinguish

based on a suspect or quasi-suspect classification. And the goal of protecting children is both legitimate and "compelling." *New York v. Ferber*, 458 U.S. 747, 756-57 (1982).

B. The Act reasonably advances its important goal of protecting children from risky drug treatments.

1. The Endocrine Society guidelines and WPATH standards of care lack evidence-based support.

Rejecting Tennessee's concerns about these uses of puberty blockers and cross-sex hormones, the district court invoked "the WPATH and Endocrine Society guidelines" to preliminarily enjoin the Act. *Skrmetti*, 2023 WL 4232308, at *22. But "optimal clinical decision making requires" support "from systematic summaries" based on high-quality evidence. Gordon Guyatt et al., *Users' Guides to the Medical Literature* 10 (McGraw Hill Education, 3rd ed. 2015). The Endocrine Society guidelines and WPATH standards lack such evidentiary support.

The GRADE method is widely accepted for rating available medical evidence. *Id.* at 16. High-quality evidence means the "true effect [of medical intervention] lies close to that of the estimate." Howard Balshem et al., *GRADE guidelines: 3. Rating the quality of evidence*, 64 J. Clinical Epidemiology 401, 404 (2011), App.461. Moderate-quality evidence means the "true effect is likely to be close to the estimate…, but there is a possibility that it is substantially different." *Id.* Low-quality evidence means the "true effect may be substantially different from the estimate."

Id. And very-low-quality evidence means the "true effect is likely to be substantially different from the estimate." *Id.*

When applied properly, the GRADE method "achieves explicit and transparent judgment" by requiring evaluators to disclose all evidence and reasons supporting their rating. Gordon Guyatt et al., *GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes*, 66 J. Clinical Epidemiology 151, 155 (2013), App.455. In general, strong recommendations should *not* be made based on low-quality evidence—only when "a panel would have a low level of regret if [later] evidence showed that their recommendation was misguided." Jeffrey C. Andrews et al., *GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength*, 66 J. Clinical Epidemiology 726, 731 (2013), App.490.

The Endocrine Society guidelines are not evidence-based. They lack support from systematic evidentiary reviews on key questions, including whether the recommended treatments ease gender dysphoria, improve mental health, affect brain development, or impact fertility. Hembree, *supra*, at 3873, App.720; Jennifer Block, *Gender dysphoria in young people is rising—and so is professional disagreement* 2-3, BMJ (2023), App.496-97. The authors did not systematically list the evidence supporting their recommendations or justify their evidence ratings. Hembree, *supra*, at 3881-83, App.728-30. They alarmingly made strong recommendations based on low-quality evidence without saying whether or why

they believe those recommendations satisfy GRADE criteria. *Id.*; Block, *supra*, at 2-3, App.496-97.

Exemplifying this problem, one co-author acknowledged that the Endocrine Society had no data—"none"—to support Guideline 2.5, which suggests "there may be compelling reasons to start cross-sex hormones prior to age 16" when treating gender dysphoria. Icahn Sch. of Med., *State of the Art: Transgender Hormone Care* at 5:38-6:18, YouTube (Feb. 15, 2019), https://www.youtube.com/watch?v=m7Xg9gZS_hg; Hembree, *supra*, at 3871, App.718. This change, he said, gave doctors "cover" to provide cross-sex hormones to children. *State of the Art* at 5:38-6:18. Such disregard supports his earlier boast that most in "the medical world [are] more conservative than [endocrinologists]." *Id.* at 4:33-4:38. So as one developer of evidence-based medicine has said, the Endocrine Society's guidelines have "serious problems." Block, *supra*, at 2, App.496.

Likewise, WPATH standards lack evidence-based support. The group admits its standards lack support from systematic reviews of available evidence and so do not rate the quality of its evidence. E. Coleman et al., *Standards of Care for the Health of Transgender and Gender Diverse People, Version 8*, 23 Int'l J. Transgender Health S1, S42 (2022), https://www.wpath.org/publications/soc. Per WPATH, "a systematic review" of the evidence "is not possible," but a co-developer of evidence-based medicine says such reviews "are always possible," and WPATH

would "violat[e] standards of trustworthy guidelines" by making "a recommendation without one." Block, *supra*, at 3, App.497. It turns out that others *have* systematically reviewed the evidence, and as detailed below, the results are disturbing.

2. Systematic reviews have shown insufficient evidence to use puberty blockers and cross-sex hormones to treat minors with gender dysphoria.

Many organizations, including the U.K. National Institute for Health & Care Excellence, have systematically reviewed available evidence supporting the use of hormonal intervention to treat gender-dysphoric minors and concluded it has "very low" quality under the GRADE method. *Evidence review: Gonadotrophin releasing hormone analogues for children & adolescents with gender dysphoria*, NICE (2020) (NICE I), App.307-437; *Evidence review: Gender-affirming hormones for children & adolescents with gender dysphoria*, NICE (2020) (NICE II), App.151-306. On this basis, England's National Health Service has *stopped* using puberty blockers to treat gender-dysphoric youth in clinical settings. *Implementing advice from the Cass Review*, NHS (2023), https://perma.cc/L2CV-M7ND.

Swedish and Finnish authorities have also systematically reviewed the evidence and concluded that its quality is insufficient to justify using puberty blockers and cross-sex hormones for children with gender dys-

phoria in clinical settings. *Medical treatment methods for dysphoria associated with variations in gender identity in minors – recommendation* 1, Council for Choices in Health Care in Finland (2020), App.537; *Care of children & adolescents with gender dysphoria* 4, Socialstyrelsen (2022), App.57. To be sure, European nations that forbid clinical use still allow research to continue. But that does *not* mean drug intervention is safe—clinical research aims to benefit future patients, not those being studied. *Clinical Research Versus Medical Treatment*, FDA (2018), https://perma.cc/8TTD-2HTP.

Likewise, McMaster University, where evidence-based medicine originated, systematically reviewed the "[e]ffects of gender affirming therapies in people with gender dysphoria" and concluded that (1) "there is great uncertainty about the effects of puberty blockers, cross-sex hormones, and surgeries in young people with gender dysphoria" and (2) available evidence "is not sufficient to support … using these treatments." Romina Brignardello-Petersen & Wojtek Wiercioch, *Effects of gender affirming therapies in people with gender dysphoria: evaluation of the best available evidence* 5 (2022), App.623. The Cochrane Library agrees, finding *not a single study* sufficiently rigorous to warrant inclusion in its systematic review. C. Haupt et al., Cochrane Library, *Anti-androgen or estradiol treatment or both during hormone therapy in transitioning transgender women (Review)* (2020), App.26-47.

And just this month, 21 clinicians and researchers from nine countries publicly warned that treating gender-dysphoric minors with puberty blockers and cross-sex hormones "is not supported by the best available evidence," expressly criticizing "the Endocrine Society's claims" to the contrary. Riittakerttu Kaltiala et al., *Youth Gender Transition is Pushed Without Evidence*, Wall St. J., July 13, 2023, https://perma.cc/5P6X-KNHL. Per this report, "[e]very systematic review of evidence to date, *including one published in the Journal of the Endocrine Society*, has found the evidence for mental-health benefits of hormonal interventions for minors to be of low or very low certainty." *Id.* (emphasis added). "By contrast, the risks are significant and include sterility, lifelong dependence on medication and the anguish of regret." *Id.*

3. Using puberty blockers and cross-sex hormones to treat minors with gender dysphoria has no proven benefits and poses substantial risk.

Bypassing these concerns, the court below downplayed the risks of using puberty blockers and cross-sex hormones to treat gender dysphoria and said those interventions might lower "rates of depression, suicide, and additional mental health issues." *Skrmetti*, 2023 WL 4232308, at *28. Yet those drugs have not proven these benefits and in fact pose substantial risk.

Start with supposed benefits. No reliable evidence suggests that drug use reduces the risk of suicide. WPATH's own commissioned review

shows no link between the use of cross-sex hormones and decreased suicide rates in gender-dysphoric individuals. Kellan E. Baker et al., *Hormone Therapy, Mental Health, & Quality of Life Among Transgender People: A Systematic Review*, 5:4 J. Endocrine Soc'y 1, 12 (2021), App.511. Multiple studies have also found high suicide rates before, during, and after attempted gender transition. C.M. Wiepjas et al., *Trends in suicide death risk in transgender people: results from the Amsterdam Cohort of Gender Dysphoria study (1972-2017)*, 141 Acta Psychiatrica Scandinavica 486, 490 (2020), App.52; Jay McNeil et al., *Suicide in Trans Populations: A Systematic Review of Prevalence and Correlates*, 4:3 Psychology of Sexual Orientation & Gender Diversity 341, 348 (2017), App.479; Cecilia Dhejne et al., *Long-Term Follow-Up of Transsexual Persons Undergoing Sex Reassignment Surgery: Cohort Study in Sweden*, 6:2 PLOS ONE 1, 5 (2011), App.64. And more alarmingly, a recent study found that rates of suicidal ideation, suicide attempts, and non-suicidal self-harm *increased* after minors began using puberty blockers and cross-sex hormones, Laura E. Kuper et al., *Body Dissatisfaction & Mental Health Outcomes of Youth on Gender-Affirming Hormone Therapy*, 145:4 Pediatrics 1, 8 (2020), App.533.

Likewise, no reliable evidence shows that drug use improves psychosocial outcomes. As the NICE systematic review found, studies showing puberty blockers and cross-sex hormones' effect on mental health outcomes trigger "very low certainty" and suggest little or no change. NICE

I, *supra*, at 13, App.319; NICE II, *supra*, at 50, App.200. Indeed, many studies report no mental health improvement after such intervention. Riittakerttu Kaltiala et al., *Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria*, 74:3 Nordic J. Psychiatry 213, 217 (2020), App.607; Annette L. Cantu et al., *Changes in Anxiety & Depression from Intake to First Follow-Up Among Transgender Youth in a Pediatric Endocrinology Clinic*, 5:3 Transgender Health 196, 198 (2020), App.19; Polly Carmichael et al., *Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK*, 16:2 PLOS ONE 1 (2021), App.576-601; Elizabeth Hisle-Gorman et al., *Mental Healthcare Utilization of Transgender Youth Before & After Affirming Treatment*, 18 J. Sexual Med. 1444, 1447 (2021), App.122.

Moving to risks, drug intervention may impair cognitive development. Researchers know that "the pubertal and adolescent period is associated with profound neurodevelopment," which depends heavily on sex-specific hormones; and many academics worry that "pubertal suppression may prevent key aspects of development during a sensitive period of brain organization." Diane Chen et al., *Consensus Parameter: Research Methodologies to Evaluate Neurodevelopmental Effects of Pubertal Suppression in Transgender Youth*, 5:4 Transgender Health 246, 248-249 (2020), App.72-73. So a respected research group published a "consensus

parameter" requesting more research on this issue, *id.*—a point supported by other reviews and reports support—and noting the critical information deficit. NICE I, *supra*, at 38, App.344; Cass Review 38-39.

Next, these drug uses may increase infertility risk. The Endocrine Society itself admits this. Hembree, *supra*, at 3878, App.725. Children who persist through their guidelines and take cross-sex hormones in early to mid-adolescence will lack "fertility preservation" options because they never develop fertility. Dep. of Armand H. Antommaria at 207:16-209:23, *Boe v. Marshall*, No. 2:22-cv-184-LCB (M.D. Ala. Apr. 21, 2023), App.751-52. While the court below cited Dr. Adkins to minimize this risk, *Skrmetti*, 2023 WL 4232308, at *24, she cites studies involving (1) two females who underwent egg preservation as adults after experiencing endogenous female puberty and later had successful surrogate pregnancies, Susan Maxwell et al., *Pregnancy Outcomes After Fertility Preservation in Transgender Men*, 129:6 Obstetrics & Gynecology 1031 (2017), App.712-15, (2) two similar females whose eggs remain preserved, Brett A. Stark & Evelyn Mok-Lin, *Fertility preservation in transgender men without discontinuation of testosterone*, 3:2 Fertil Steril Rep 153 (2022), App.22-25, and (3) a non-representative survey of females who self-reported pregnancy after taking testosterone as adults, Alexis D. Light et al., *Transgender Men Who Experienced Pregnancy After Female-to-Male Gender Transitioning*, 124:6 Obstetrics & Gynecology 1120, 1126 (2014),

App.7. None of these studies considered representative samples, individuals who received puberty blockers or cross-sex hormones as adolescents, or males treated with estrogen. Such limited studies hardly resolve Tennessee's reasonable concerns.

These drug interventions may also weaken bone density. For adults, osteoporosis is a "well understood" risk of using cross-sex hormones long-term. Cass Review 36. And children face added risks. Because bone mineral density increases during puberty, children undergoing puberty suppression do not experience this full increase. Hembree, *supra*, at 3882, App.729. And evidence suggests these children never catch up. *Id.*; NICE II, *supra*, at 14, App.164. While the court below noted evidence about children treated with puberty suppression for central precocious puberty who later suffered "no changes in bone mineralization," *Skrmetti*, 2023 WL 4232308, at *25, these children stopped puberty suppression and underwent endogenous puberty consistent with their natal sex. R.113-4, 1295-99. Evidence that endogenous puberty allowed these children to achieve normal adult bone mineral density says nothing about whether children who never undergo endogenous puberty will have the same result. *Id.*

Concerns also exist about cardiovascular health. The Endocrine Society admits evidence shows that cross-sex hormones detrimentally affect adult lipid profiles. Hembree, *supra*, at 3891, App.738. This is a "well understood" risk. Cass Review 36. NICE's systematic review uncovered

only one cardiovascular study of individuals who began cross-sex hormones in adolescence, and it found statistically significant increases in blood pressure and body mass for both sexes and worsening lipid profiles for natal females. NICE II, *supra*, at 14, App.164. Both the Endocrine Society and NICE say we need better studies to show how long-term use of cross-sex hormones beginning in adolescence affects cardiovascular health. Hembree, *supra*, at 3891, App.738; NICE II, *supra*, at 14, App.164.

These drug interventions may also limit sexual function. WPATH's president has reported that "about zero" natal males can achieve orgasm after undergoing early puberty suppression followed by cross-sex hormones and vaginoplasty. Michael Biggs, *The Dutch Protocol for Juvenile Transsexuals: Origins & Evidence*, J. Sex & Marital Therapy 12-13 (2022), App.566-67. While this issue needs more study, *id.*, there are substantial concerns with subjecting prepubertal children to medical treatment that may affect lifelong sexual function in ways they cannot possibly understand. Stephen B. Levine et al., *Reconsidering Informed Consent for Trans-Identified Children, Adolescents, & Young Adults*, J. Sex & Marital Therapy 15 (2022), App.704.

What's more, the long-term safety of "treatments in children and adolescents with gender dysphoria" is "largely unknown" because many identified risks tend to manifest later in life—e.g. the risk of cognitive impairment, cardiovascular decline, and osteoporosis. NICE II, *supra*, at

14, App.164. Indeed, early studies report substantial *increases* in mortality from suicide, cardiovascular events, and other problems more than ten years after drug and surgical intervention. One study found that suicide rates surged *over 19 times* the rate of controls in this population, and that mortality rates from cardiovascular disease more than doubled. Dhejne, *supra*, at 5, App.64. Another study found that adults treated with cross-sex hormones faced increased long-term risk of death by suicide, stroke, and ischemic heart disease. Henk Asscheman et al., *A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones*, 164:4 Eur. J. Endocrinology 635, 635-42 (2011).

4. Drug companies have not sought regulatory approval for puberty blockers and cross-sex hormones to treat minors with gender dysphoria.

Given these concerns, it's no surprise that drug companies have not sought FDA approval to treat gender-dysphoric minors with hormonal interventions. Under federal law, a pharmaceutical company wanting to introduce any new drug into commerce must first obtain FDA approval. 21 U.S.C. § 355(a). If the FDA finds that the drug is safe and effective for use under conditions prescribed in proposed labeling, the pharmaceutical company can introduce the new drug into commerce using the approved labeling. *Id.* § 355(d). If the company seeks to modify its labeling to add a new use, the company must submit a new drug application seeking

FDA approval for the change under the same process as the initial approval. 21 U.S.C. § 355(b); 21 C.F.R. §§ 314.54, 314.70.

Because the FDA typically limits its review to the proposed labeling, the agency does not evaluate the safety of a new drug for off-label (i.e., unapproved) uses. *Understanding Unapproved Use of Drugs "Off Label,"* U.S. Food & Drug Admin. (2018), https://perma.cc/Y5LE-S9PZ. While clinicians may prescribe approved drugs for off-label uses when they believe it's "medically appropriate for their patient," *id.*, a drug manufacturer may not promote off-label uses of its drug. *See* 21 C.F.R. § 202.1(e)(4). In fact, many manufacturers have faced significant criminal and civil penalties for unlawfully promoting off-label uses. *E.g.*, Press Release, U.S. Dep't Justice, *Pfizer to Pay $2.3 Billion for Fraudulent Marketing* (Sept. 2, 2009), https://perma.cc/W3JG-WPBE; Press Release, U.S. Atty's Off., *Abbott Laboratories and AbbVie Inc. to Pay $25 Million to Resolve False Claims Act Allegations of Kickbacks and Off-Label Marketing of the Drug TriCor* (Oct. 26, 2018), https://perma.cc/46HZ-CEPD; Press Release, U.S. Dep't of Justice, *Endo Pharmaceuticals and Endo Health Solutions to Pay $192.7 Million to Resolve Criminal and Civil Liability Relating to Marketing of Prescription Drug Lidoderm for Unapproved Uses* (Feb. 21, 2014), https://perma.cc/56G5-NH7F.

So pharmaceutical companies must decide whether it makes financial sense to seek FDA approval for off-label use. Such applications must present ample evidence showing that the drug is safe and effective under

the proposed labeling. 21 U.S.C. § 355(d). This effort may cost companies substantial time and investment and, importantly, may reveal significant safety concerns with the new labeling. Often, companies lack financial incentive to seek such approval. That's true for AbbVie, Inc, manufacturer of the puberty blocker Lupron, which netted $783 million from sales in 2021, *Financial Release*, AbbVie (2021), App.438-50, and for Endo Pharmaceuticals, manufacturer of the puberty blocker Supprelin, which netted over $114 million from sales the same year. *Endo Reports Fourth-Quarter & Full-Year 2021 Financial Results*, ENDO (2022), App.130-50.

With these massive profits and little scientific support, drug companies have no incentive to seek FDA approval for using puberty blockers and cross-sex hormones to treat gender dysphoria. In fact, Endo has said it "has no plans to seek regulatory approval for the use of its drug for" this purpose. Chad Terhune et al., *As more transgender children seek medical care, families confront many unknowns*, Reuters (Oct. 6, 2022), https://perma.cc/UYT2-GEHC. Without regulatory approval, the Minors Protection Act safeguards Tennessee children from becoming human experiments for off-label drug use that the FDA has never approved.

5. That puberty blockers and cross-sex hormones are used to treat different physical illness does not make them safe to treat gender dysphoria.

The district court criticized the Act for allowing these drugs to treat medical issues like "congenital defect[s], precocious puberty," or "physical injury" but forbidding them to treat gender dysphoria. *Skrmetti*, 2023 WL 4232308, at *30. This distinction is not "arbitrary." *Id.* The Act validly distinguishes what's safe to treat mental health conditions from what's safe to treat physical conditions that have different etiologies, diagnostic criteria, and treatment pathways. Ignoring such distinctions would allow litigants to argue it's safe to prescribe chemotherapy drugs to treat anxiety because doctors use them to treat cancer.

Substantial differences separate different uses of puberty blockers and cross-sex hormones. For example, central precocious puberty occurs when a child experiences puberty earlier than normal. It is diagnosed through physical examination and laboratory testing. Melinda Chen & Erica A. Eugster, *Central Precocious Puberty: Update on Diagnosis & Treatment*, 17:4 Paediatr Drugs 273, 275 (2015), App.541. And it is treated through puberty blockers, though the patient stops these drugs in time to undergo endogenous puberty. By contrast, puberty blockers are administered for gender dysphoria *during* the normal ages for puberty, and when stopped, the child is prescribed cross-sex hormones to avoid endogenous puberty altogether. R.113-4, 1295-99, 1314.

Likewise, polycystic ovary syndrome occurs when females overproduce testosterone. It is diagnosed through observation, imaging, and laboratory testing, and patients are often treated with estrogen to suppress testosterone, which aims to counteract the ill effects of abnormal hormone levels. A recent study showed that treating this condition with testosterone suppression may preserve fertility otherwise impaired by abnormal testosterone levels. E. Elenis et al., *Early initiation of anti-androgen treatment is associated with increased probability of spontaneous conception leading to childbirth in women with polycystic ovary syndrome: a population-based multiregistry cohort study in Sweden*, 36:5 Human Reproduction 1427, 1433-34 (2021), App.116-17. Yet testosterone blockers for gender dysphoria thwart fertility and normal sexual function, as explained above.

Finally, consider sexual development disorders, which expert witness Dr. Hruz routinely treats at the Washington University clinical program he co-founded. All these disorders involve objective chromosomal or physical abnormalities. R.113-4, 1288-89. When used in these situations, drug intervention helps physically unhealthy individuals develop physically healthy sexual function consistent with their sex. In stark contrast, using puberty blockers and cross-sex hormones for gender dysphoria causes physically healthy individuals to *lose* healthy sexual function consistent with their sex. That's a critical difference, which justifies Tennessee's Minors Protection Act.

6. Consistent with evidence-based medicine, this Court should credit experts based mainly on their analysis of published evidence.

The court below "discount[ed]" Tennessee's experts because they did not report significant experience treating minors with gender dysphoria and had been falsely accused of bias. *Skrmetti*, 2023 WL 4232308, at *20. But evidence-based medicine relies not on an expert's treating experience but on "the best available … evidence from systematic research." David L. Sackett et al., *Evidence based medicine: what it is and what it isn't*, 312 BMJ 71, 71 (1996), App.68; Guyatt, *Users' Guides*, *supra*, at xxiv. Systematic reviews rate the quality of evidence available—placing randomized controlled trials at the top, and "unsystematic observations of individual clinicians" at the bottom. *Id.* at 15-16. Yet the court below valued the weakest evidence the most and, in so doing, minimized Dr. Cantor's and Dr. Hruz's evidence-based reports.

Dr. Cantor is a Ph.D. clinical psychologist and researcher with decades of experience and dozens of peer-reviewed publications. He supported his 312-paragraph expert report with citations to 253 different sources, most of which were peer-reviewed medical literature, including his own peer-reviewed publications about treating gender dysphoria.

Dr. Hruz holds an M.D. and a Ph.D. in biochemistry. He's a board-certified pediatrician and pediatric endocrinologist with decades of clinical experience, a medical faculty appointment, and dozens of peer-reviewed publications. He supported his 152-paragraph expert report with

344 footnoted citations to medical literature. Dr. Hruz also regularly prescribes puberty blockers and cross-sex hormone, consistent with evidence-based medicine, to treat endocrinal disorders. R.113-4, 1285-86.

Despite all this, the court below found Dr. Cantor's and Dr. Hruz's expert reports "minimally persuasive" because they did not say that they have "diagnosed or treated a minor with gender dysphoria." *Skrmetti*, 2023 WL 4232308, at *20. That preference for treatment experience defies evidence-based medicine.

The court below also wrongly devalued Dr. Levine's and Dr. Laidlaw's expert reports. Dr. Levine is a clinical psychiatrist with almost 50 years of experience treating gender dysphoria, a tenured professor, the author of over 180 peer-reviewed publications, and the chairperson of the committee that developed the fifth version of WPATH's standards of care. He submitted a 238-paragraph expert report below citing more than 100 scientific sources.

Dr. Laidlaw is a board-certified endocrinologist who's authored multiple peer-reviewed publications and has over 20 years of experience treating patients with hormonal disorders. He submitted a 304-paragraph expert report citing more and 100 scientific sources.

Though Dr. Levine and Dr. Laidlaw both provided superior evidence-based testimony, the district court criticized their opinions con-

cerning the effects of these drug interventions on (1) fertility, (2) cardiovascular health, (3) sexual response, (4) bone mineral density, and (5) cancer risks. But the best studies support their reports.

Start with infertility risk. Dr. Levine quoted Endocrine Society guidelines to support his report, (R.113-5, 1460, Hembree, *supra*, at 3880, App.727), and Dr. Laidlaw cited Dutch studies showing that children who begin puberty suppression almost always progress to surgical sterilization, (R.113-7, 1562). But Plaintiffs' own experts and the district court relied on Endocrine Society guidelines. And Dr. Laidlaw's concern is based on Dutch research, which is trusted by those who promote drugs and surgeries for children with gender dysphoria. So there is nothing to discredit here.

The district court trusted anecdote and largely off-topic studies instead. For example, it cited one eight-year study showing that four months after stopping testosterone treatment, some females identifying as male had comparable egg yields to other females, but the study was admittedly limited by its "selection of patients with infertility as the comparison group." Angela Leung et al., *Assisted reproductive technology outcomes in female-to-male transgender patients compared with cisgender patients: a new frontier in reproductive medicine*, 112:5 Fertility & Sterility 858, 859 (2019), App.10. There was no control group without congenital fertility problems.

Dr. Levine also cited multiple peer-reviewed studies showing a link between cross-sex hormones and cardiovascular risk. R.113-5, 1461. He then agreed with the NICE systematic review showing that this risk has been insufficiently studied and that long-term term studies are needed because cardiovascular risks develop over time and often manifest only after many years. R.113-5, 1461-62. Because Dr. Levine opined that more research was needed, the district court dismissed potential cardiovascular risk as "speculative," and instead credited Dr. Adkins's anecdotal observation that her patients have suffered no short-term cardiovascular decline. *Skrmetti*, 2023 WL 4232308, at *26. This finding overlooks Dr. Levine's main concern—that *long-term* cardiovascular risks should at least be studied *before* injecting minors with potentially dangerous drugs.

Dr. Levine next cited peer-reviewed literature to support concerns about puberty suppression's effect on sexual response—concerns that track his vast clinical experience. R.113-5, 1460 (citing Levine (2018)). The district court mistakenly rejected that testimony by citing Endocrine Society guidelines on sexual response following genital surgery (not puberty suppression) and WPATH guidelines addressing adults (not children undergoing suppression)—two issues Dr. Levine was not addressing.

Dr. Levine also cited four peer-reviewed studies showing puberty suppression's risk to bone density. R.113-5, 1456. He then noted conflicting evidence on this issue before concluding that such treatment cannot

be considered safe given the uncertainty. *Id.* Dr. Adkins does not dispute that evidence suggests puberty suppression prevents normal bone density development or that treated minors may not fully catch up once they take cross-sex hormones. R.141, 2392 Instead, she marshals evidence from much younger children treated for precocious puberty who experience no bone density deficits, which as explained above, proves nothing because *endogenous puberty* allowed these children to achieve normal bone mineral density as adults. Section I.B.3, *supra*. The court overlooked this error, and never even addressed Dr. Laidlaw's concerns on this topic. R.113-7, 1564-66.

Finally, Dr. Laidlaw cited Endocrine Society guidelines and peer-reviewed literature to show that using cross-sex hormones may increase cancer risk. R.113-7, 1572-73, 1576. Rather than crediting this literature, the district court relied on Dr. Adkins's assertion that she has "rarely seen" cancer develop in her clinical practice. *Skrmetti*, 2023 WL 4232308, at *27. That's exactly the kind of unscientific observation that should not be prioritized over a risk identified in peer-reviewed literature.

The district court also noted that Endocrine Society guidelines say breast augmentation and mastectomy do not appear to increase cancer risk. *Id.* But Dr. Laidlaw testified that *cross-sex hormones* increase cancer risk, a concern that the Endocrine Society itself acknowledges.

In sum, the court below misvalued, misinterpreted, and misapplied the scientific evidence. Those errors led the court to mistakenly hold the

Minors Protection Act unconstitutional. The errors also led the court to promote an experimental treatment that European countries—the world leaders in treating gender dysphoria—have concluded is inconclusive at best and harmful at worst. American kids deserve better.

II. This Court should allow state legislatures to decide this difficult medical issue rife with uncertainty and so avoid miring courts further in constitutionalized medicine.

"It is indisputable 'that a State's interest in safeguarding the physical and psychological wellbeing of a minor is compelling.'" *Otto v. City of Boca Raton*, 981 F.3d 854, 868 (11th Cir. 2020) (quoting *Ferber*, 458 U.S. at 756-57). And States play a "significant role … in regulating the medical profession." *Gonzales v. Carhart*, 550 U.S. 124, 157 (2007). Here, Tennessee has enacted the Minors Protection Act to safeguard children from potentially dangerous and experimental drug treatments. Section I, *supra*. Evidence strongly suggests that Tennessee's caution is warranted. *Id*. But even if both sides had "medical support for their position," "[m]edical uncertainty does not foreclose the exercise of legislative power," and the State may reasonably act to protect children. *Gonzales*, 550 U.S. at 161, 164.

What's more, both sides have marshaled experts to support their positions. These experts belong to professional groups, but "their institutional positions cannot define the boundaries of" what the Constitution requires. *Otto*, 981 F.3d at 869. "They may hit the right mark," or they

may "miss it." *Id.* And sometimes, these professional communities can be wrong "by a wide margin." *Id*. Indeed, it's "not uncommon for professional organizations to do an about-face in response to new evidence or new attitudes." *Id.* That's happened on the very issue presented here, as European nations are now backtracking and forbidding these drug interventions to treat children with gender dysphoria because new evidence suggests that caution is best. Section I.B.2, *supra*.

For this reason, courts give "state and federal legislatures wide discretion to pass legislation in areas where there is medical and scientific uncertainty." *Gonzales*, 550 U.S. at 163. This restraint is both wise and constitutionally required. Take *Roe v. Wade*, 410 U.S. 113 (1973), a case in which the Court constitutionalized abortion without textual support or certainty about unborn human life. Courts then struggled for decades to apply an "inherently standardless" rule covering an issue "of great social significance." *Dobbs v. Jackson Women's Health Org.*, 142 S. Ct. 2228, 2272, 2284 (2022). Then just last year, the Court reversed *Roe*, admitting that precedent had "departed from [the Court's] normal rule" of legislative deference and regretting the tremendous "turmoil" that deviation inflicted. *Id.* at 2268, 2283. This Court should avoid similar turmoil by deferring to reasonable legislative judgment here.

CONCLUSION

This Court should reverse and uphold Tennessee's right to protect children consistent with its reasonable legislative judgment. What's more, the Court should explicitly condemn the district court's rejection of evidence-based medicine and experts because that court's pronouncements might mistakenly lead families to authorize experimental treatments that could result in permanent harm to their children.

Dated: July 24, 2023

Respectfully submitted,

/s/ John J. Bursch

JACOB P. WARNER
ALLIANCE DEFENDING FREEDOM
15100 N. 90th Street
Scottsdale, AZ 85260
(480) 444-0020
jwarner@ADFlegal.org

JOHN J. BURSCH
JAMES A. CAMPBELL
ALLIANCE DEFENDING FREEDOM
440 First Street NW
Suite 600
Washington, DC 20001
(616) 450-4235
jbursch@ADFlegal.org

Attorneys for Amicus Curiae

RULE 32(G)(1) CERTIFICATE OF COMPLIANCE

This brief complies with the word limit of Fed. R. App. P. 32(a)(7)(B) and 29(a)(5) because this brief contains 6,496 words, excluding parts of the brief exempted by Fed. R. App. P. 32(f) and 6th Cir. R. 32(b).

This brief complies with the typeface requirements of Fed. R. App. P. 32(a)(5) and the type-style requirements of Fed. R. App. P. 32(a)(6) because this brief has been prepared in Word 365 using a proportionally spaced typeface, 14-point Century Schoolbook.

Dated: July 24, 2023

/s/ John J. Bursch
John J. Bursch

Attorney for Amicus Curiae

CERTIFICATE OF SERVICE

I hereby certify that on July 24, 2023, I electronically filed the foregoing brief with the Clerk of the Court for the United States Court of Appeals for the Sixth Circuit by using the appellate CM/ECF system. I certify that all participants in the case are registered CM/ECF users and that service will be accomplished by the CM/ECF system.

/s/ John J. Bursch
John J. Bursch

Attorney for Amicus Curiae

APPENDIX

Appendix Table of Contents

Original Research

Transgender Men Who Experienced Pregnancy After Female-to-Male Gender Transitioning

Alexis D. Light, MD, MPH, Juno Obedin-Maliver, MD, MPH, Jae M. Sevelius, PhD, and Jennifer L. Kerns, MD, MPH

OBJECTIVE: To conduct a cross-sectional study of transgender men who had been pregnant and delivered after transitioning from female-to-male gender to help guide practice and further investigation.

MATERIALS AND METHODS: We administered a web-based survey from March to December 2013 to inquire about demographics, hormone use, fertility, pregnancy experience, and birth outcomes. Participants were not required to have been on hormone therapy to be eligible. We used a mixed-methods approach to evaluate the quantitative and qualitative data.

RESULTS: Forty-one self-described transgender men completed the survey. Before pregnancy, 61% (n=25) had used testosterone. Mean age at conception was 28 years with a standard deviation of 6.8 years. Eighty-eight percent of oocytes (n=36) came from participants' own ovaries. Half of the participants received prenatal care from a physician and 78% delivered in a hospital. Qualitative themes included low levels of health care provider awareness and knowledge about the unique needs of pregnant transgender men as well as a desire for resources to support transgender men through their pregnancy.

CONCLUSION: Transgender men are achieving pregnancy after having socially, medically, or both transitioned. Themes from this study can be used to develop transgender-appropriate services and interventions that may improve the health and health care experiences of transgender men.

(Obstet Gynecol 2014;124:1120–7)
DOI: 10.1097/AOG.0000000000000540

Transgender individuals often report many barriers in attempting to access health care.[1] The American College of Obstetricians and Gynecologists (the College) recently called on obstetrician–gynecologists to help eliminate these barriers for transgender men (also called female-to-male individuals) by creating nondiscriminatory practices, assisting with gender transition, and providing transgender-appropriate and comprehensive health care.[2] Despite the College's call to action, little systematic attention has been paid to the health and reproductive experiences of transgender men or those individuals who are born with female sexual organs but who identify as male.

Transgender men are individuals who have a male or masculine gender identity but were assigned female at birth. The gender affirmation process may include social, medical, and surgical aspects of transition, although not all transgender men desire medical intervention.[3] Many transgender men desire children[4] and there are anecdotal reports supporting the biological possibility of pregnancy for transgender men who retain a uterus and discontinue testosterone therapy.[5–7] However, there is little scientific literature describing pregnancy experiences among transgender men or the effects of exogenous administration of testosterone on fertility, pregnancy, and neonatal outcomes.[8] Understanding transgender men's experiences with fertility, pregnancy, and birth will allow health care providers to augment pre- and posttransition discussions regarding fertility options, the roles of cross-sex hormones on fecundity, potential birth outcomes, and to support their physical and mental well-being during pregnancy. Expanded knowledge may also help

From the Washington Hospital Center Department of Obstetrics and Gynecology, Washington, DC; the Department of Obstetrics, Gynecology, and Reproductive Sciences, University of California, San Francisco, and the University of California, San Francisco Center of Excellence for Transgender Health, San Francisco, California.

Corresponding author: Alexis D. Light, MD, MPH, 2360 Champlain Street, NW, Apartment 2.1, Washington, DC 20009; e-mail: Alexis.D.Light@gunet.georgetown.edu.

Financial Disclosure
The authors did not report any potential conflicts of interest.

ISSN: 0029-7844/14

health care providers support transgender men in attaining and maintaining healthy pregnancies.

We conducted a mixed-methods study to explore the experiences of transgender men and to contribute to the knowledge base of fertility, conception, pregnancy experience, and birth outcomes among transgender men.

MATERIALS AND METHODS

We conducted a cross-sectional survey from March to December 2013 of transgender men (assigned female at birth with a masculine, transmasculine, transmale, or female-to-male gender identity) who had been pregnant and delivered a neonate. Inclusion criteria were: age older than 18 years, self-identification as male before pregnancy, pregnancy within the last 10 years, and the ability to fill out the survey in English. Eligibility criteria did not require any type of medical (eg, testosterone use) or surgical (eg, bilateral mastectomy) transition. We recruited study participants through convenience sampling and we collected data using a web-based survey. Participation was not limited by geographic location.

We administered the online survey through REDCap,[9] an encrypted and secure online survey platform. The study contained 47 multiple-choice questions and 24 questions addressing demographics, hormone use, fertility, pregnancy experience, birth experience, and fetal outcomes. The survey concluded with four open-ended questions: "Is there anything you would like medical providers to know about transgender men and pregnancy?" "What was the experience of being pregnant like for you?" "What was the experience of giving birth like for you?" "What was the postpartum experience like for you?" The survey was developed by the authors in consultation with the Center of Excellence for Transgender Health at University of California, San Francisco and other health care providers serving the transgender community.

Initial recruitment occurred through distribution to key stakeholders in lesbian, gay, bisexual, and transgender health centers; transgender community groups; and Internet-based social networking pages created by study authors. We recruited additional participants through initial contacts. We provided interested individuals with a comprehensive study description and links to the study. After accessing the electronic study web site, participants were presented with informed consent documents and participants confirmed their consent through accessing a link to web-based survey. No in-person contact was made with survey participants.

We conducted a mixed-methods analysis to evaluate the quantitative and qualitative data collected from the survey. Using STATA 13.0, we performed unadjusted analyses using χ^2 for method of delivery; *t* tests for pregnancy age, body mass index, and gestational age; and Fisher's exact for all other variables according to testosterone use before pregnancy. As a result of nonresponse, variable totals may not sum to column totals or within category totals. A *P* value of $\leq .05$ was considered statistically significant. We analyzed the qualitative data using grounded theory, identifying iterative themes, and adding new codes as concepts emerged.[10] This study was approved by the University of California, San Francisco Committee on Human Research.

RESULTS

We excluded nine of the 56 participants who began the survey as a result of insufficient responses for analysis, and six others were excluded because they did not meet study criteria indicating male gender before pregnancy.[11] We included participants who identified as female or preferred "she" or "her" pronouns only if they had more than one validating indicator of a transgender identity (use of testosterone, male identity with female pronouns, or female identity with male pronouns). Forty-one participants remained for final analysis (Table 1). Most of our participants were from the western United States, identified as white, and had completed at least some college. Pronoun preference differed between those who had used testosterone and those who had not ($P=.04$). Participants who had previously used testosterone were more likely to prefer the pronoun "he," whereas those who had not used testosterone were more likely to identify with "they." Although most respondents were primiparous, those who had not used testosterone were more likely to be multiparous ($P=.006$). Four transgender men (10%), all of whom had been on testosterone previously, reported a prior diagnosis of polycystic ovary syndrome.

Twenty-five (61%) transgender men reported using testosterone before pregnancy (Table 2). Among those who had used testosterone, 20 (80%) reported resuming menstruation within 6 months after stopping testosterone. Five participants (20%) conceived while still amenorrheic from testosterone use. After pregnancy, six (38%) participants who had not previously used testosterone before pregnancy initiated use. Ten participants (40%) who had been on previously testosterone reported that they had not yet resumed testosterone use after pregnancy.

Table 1. Participant Characteristics

Characteristic	All (N=41)	Prior Testosterone Use		P
		Yes (n=25)	No (n=16)	
Age (y)*	28±6.8	29±6.9	27±6.8	.5
Gender identity†				.07
Male	21 (51)	12 (48)	9 (56)	
Transgender, female-to-male, transman	10 (24)	9 (36)	1 (6)	
Bigender, gender fluid, genderqueer	8 (20)	3 (12)	5 (31)	
Female	1 (2)	1 (4)	0	
Other	1 (2)	0	1 (6)	
Personal pronoun preference‡				.04
He	32 (82)	21 (88)	11 (73)	
They	3 (8)	0	3 (20)	
She	2 (5)	2 (8)	0	
Ey	1 (2)	1 (4)	0	
No pronouns	1 (2)	0	1 (7)	
Country				.4
United States	35 (85)	20 (80)	15 (94)	
Outside United States§	6 (15)	5 (20)	1 (6)	
U.S. region‡‖				.9
West	19 (59)	11 (61)	8 (57)	
Northeast	5 (16)	3 (17)	2 (14)	
South	5 (16)	2 (11)	3 (21)	
Midwest	3 (9)	2 (11)	1 (7)	
Race or ethnicity‡				1.0
White	36 (92)	21 (88)	15 (100)	
Asian	1 (3)	1 (4)	0	
Asian and black	1 (3)	1 (4)	0	
Native Hawaiian or other Pacific Islander	1 (3)	1 (4)	0	
Education level‡				.7
High school degree or less	4 (10)	3 (12.5)	1 (7)	
Vocational training or some college	12 (31)	6 (25)	6 (40)	
Associate or Bachelor's degree	14 (36)	10 (42)	4 (27)	
Master's or doctoral degree	9 (23)	5 (21)	4 (27)	
Annual household income ($)‡				.4
Less than 20,000	6 (15)	2 (8)	4 (25)	
20,000–59,999	20 (49)	12 (50)	8 (50)	
60,000–100,000	8 (20)	6 (25)	2 (13)	
More than 100,000	5 (13)	4 (17)	1 (7)	
Multiparous (2 or more pregnancies)	15 (37)	5 (20)	10 (63)	.006
Previous PCOS diagnosis	4 (10)	4 (16)	0	.15
BMI at the start of pregnancy (kg/m^2)	26±6	26±6	27±6	.6
Gender-confirming surgical procedure‡¶				.7
Bilateral mastectomy	19 (46)	13 (52)	6 (38)	
Oophorectomy	2 (5)	0	2 (13)	
Hysterectomy	2 (5)	2 (8)	0	
Phalloplasty or metoidioplasty#	1 (2)	1 (4)	0	

PCOS, polycystic ovary syndrome/BMI, body mass index.
Data are mean±standard deviation or n (%) unless otherwise specified.
* Age at the beginning of their most recent pregnancy.
† Kuper et al.[28]
‡ Not all the participants answered this question.
§ Canada (n=2), Germany (n=1), England (n=1), Israel (n=1), and Switzerland (n=1).
‖ Regions were defined according to the 2010 U.S. census.
¶ Surgery may have occurred before or after pregnancy.
# Metoidioplasty is procedure that separates the clitoris from the labia to assume a physiologic position similar to a penis (Djordjevic et al[29]).

Two thirds of pregnancies were planned (Table 3). Before the most recent pregnancy, condoms were the most common form of contraception followed by no form of contraception and abstinence (defined as not engaging in penile–vaginal intercourse). Those who had previously used testosterone were more likely to

Table 2. Findings Among Those Who Used Testosterone Before Pregnancy of Report (n=25)

Characteristic	Value
Age (y) when testosterone was initiated	25 (17–35)
Length of testosterone use before pregnancy (y)	
Less than 1	10 (40)
1–2	6 (24)
3–10	4 (16)
More than 10	5 (20)
Stopped taking testosterone to become pregnant	17 (68)
Duration between stopping testosterone and resumption of menses (mo)	
No menses before pregnancy	5 (20)
Less than 1	2 (8)
1	6 (24)
2	7 (28)
3	4 (16)
4–6	1 (4)
Resumed or initiated testosterone after pregnancy*	20 (48)

Data are median (range) or n (%).
* Of total respondents in the study (N=41).

report no contraceptive use or abstinence, whereas those who had not used testosterone were more likely to use a hormonal contraceptive method (P=.03). The majority of oocytes came from the participants' own ovaries, whereas the majority of sperm came from a significant other or spouse. Most transgender men became pregnant within 4 months of trying, only 15% had a preconception medical consultation, and 7% used fertility drugs to become pregnant.

Pregnancy, delivery, and birth outcomes did not differ according to prior testosterone use (Table 4). Half of the participants received prenatal care from a physician, 40% from an obstetrician, and 10% from a family medicine physician. More than three fourths of the participants began taking prenatal vitamins either before pregnancy or within the first trimester, whereas 15% reported not taking any prenatal vitamins. Participants reported a variety of perinatal complications including hypertension (12%), preterm labor (10%), placental abruption (10%), and anemia (7%). Anemia was not reported by participants who had previously used testosterone. A higher proportion of transgender men who had used testosterone underwent cesarean delivery compared with those who reported no testosterone use (36% compared with 19%, respectively), although this finding was not statistically significant. Among those who underwent a cesarean delivery, 25% cited the indication as elective. Those who had previously used testosterone were statistically less likely to chest (breast) feed their infant than those who had not previously used testosterone (P=.04).

Thirty participants (73%) answered at least one of the four open-ended questions. Major themes from these responses were: 1) effect of pregnancy on concepts of family structure; 2) isolation; 3) gender dysphoria and pregnancy; and 4) interactions with health care providers.

Many participants discussed their pregnancy in the context of family structure. For some, pregnancy was a necessary step in creating the family they desired: "I looked at it as something to endure to have a child" (36-year-old, prior testosterone use). Others described the pregnancy in pragmatic terms, possibly as a way to avoid gender dissonance: "Like my body was a workshop, building up this little kid" (35-year-old, prior testosterone use). Another participant found a way to embrace the pregnancy, describing the pregnancy and birth as a bridge to fatherhood: "Pregnancy and childbirth were very male experiences for me. When I birthed my children, I was born into fatherhood" (29-year-old, no prior testosterone use). Participants often used words such as "dad," "carrier," and "gestational parent" to affirm their male gender identity and describe their parenting role.

Feelings of isolation were common. One participant stated, "Pregnancy came with feelings of isolation and limitation" (28-year-old, prior testosterone use). Some identified the source of isolation as stemming from feeling "lonely because I was the only one" (30-year-old, prior testosterone use). These feelings were contextualized by comments about "lack of support" and "lack of resources available to pregnant transgender men." This isolation was also referenced in terms of invisibility: "I passed as 'not pregnant' until my eighth month, because I'm chubby anyways, and because people don't assume that someone who looks like me could be pregnant" (34-year-old, no prior testosterone use). As another participant simply put it: "We exist. And we are different" (35-year-old, prior testosterone use).

Another theme that emerged was the relationship between gender dysphoria and pregnancy. Some participants reported improvements in gender dysphoria, feeling new connections with their bodies: "It was relieving to feel comfortable in the body I'd been born with" (20-year-old, no prior testosterone use). Others felt an increase in dysphoria, and for some, that dysphoria continued into the postpartum period: "Heavy time, having a baby, not passing as male, all the changes and a society telling me to just be happy"

Table 3. Fertility Experiences Surrounding Most Recent Pregnancy by Prior Testosterone Use

Characteristic	Total (N=41)	Prior Testosterone Use		*P*
		Yes (n=25)	No (n=16)	
Planned pregnancy	28 (68)	19 (76)	9 (56)	.3
Contraception use before this pregnancy*†				.03
Condoms	16 (41)	10 (40)	6 (43)	
None	15 (38)	12 (48)	3 (21)	
Abstinence‡	3 (7)	3 (12)	0	
Fertility awareness	2 (8)	0	2 (14)	
Combined hormonal contraception (OCPs, transdermal patch, vaginal ring)	1 (3)	0	1 (7)	
Injection, intrauterine device, implant	1 (3)	0	1 (6)	
Partner had vasectomy	1 (3)	0	1 (6)	
Time to conception (mo)†				.14
Unplanned pregnancy	13 (32)	6 (24)	7 (44)	
Less than 1	3 (17)	1 (20)	2 (12)	
1–3	9 (22)	8 (32)	1 (6)	
4–6	8 (19)	5 (20)	3 (19)	
More than 7	4 (10)	1 (4)	3 (18)	
Source of oocyte				.12
Own ovaries	36 (88)	21 (84)	15 (94)	
Significant other or spouse	4 (10)	4 (16)	0	
Anonymous donor	1 (2)	0	1 (6)	
Source of sperm				.5
Significant other, spouse, or romantic partner	31 (76)	17 (68)	14 (88)	
Known donor	4 (10)	3 (12)	1 (6)	
Anonymous donor or sperm bank	6 (15)	5 (20)	1 (6)	
Medical intervention to become pregnant§				
Consultation	6 (15)	4 (16)	2 (12)	
Fertility drugs	3 (7)	2 (8)	1 (6)	
Assisted reproductive technology‖	5 (12)	5 (20)	0	

OCP, oral contraceptive pill.
Data are n (%) unless otherwise specified.
* Participants were given the option to identify with more than one, so total exceeds 100%.
† Not all the participants answered this question.
‡ Defined as not having penile–vaginal intercourse.
§ Participants could mark more than one, therefore not comparing the results statistically.
‖ Includes artificial insemination, in vitro fertilization, and gamete intrafallopian transfer.

(35-year-old, prior testosterone use). Combined with feelings of isolation postpartum, many participants specifically mentioned having postpartum depression. "Began to show symptoms of postpartum depression long before anyone discussed symptoms to watch for… Began researching and working through postpartum depression issues independently; found no professional with familiarity with 'trans/genderqueer' gestational parents" (28-year-old, prior testosterone use). As mentioned, the depression seemed amplified by a lack of gender-sensitive resources for postpartum depression.

In response to queries interactions with health care providers, some participants mentioned positive interactions with their health care teams regarding their gender identity. "I was always called 'he,' I was always called 'dad,' and my body parts were called by the words I used" (34-year-old, prior testosterone use). As previously, positive experiences often focused on proper use of gender-related language. Other participants mentioned negative experiences that ranged from improper pronoun use and rude treatment to being turned away from medical practices and denied treatment. In one extreme experience, a participant reported that "Child Protection Services was alerted to the fact a 'tranny' had a baby" (21-year-old, prior testosterone use). Many participants called for better treatment from the health care system through acknowledging the unique identities of pregnant transgender men and grounding health care provider–patient interactions in compassion and respect. As one participant said, "treat us as if we are normal human beings with normal bodies" (37-year-old, no prior testosterone use). Additionally, participants

Table 4. Pregnancy Experience and Neonatal Outcomes

Characteristic	Total (N=41)	Prior Testosterone Use		P
		Yes (n=25)	No (n=16)	
Source of prenatal care*				1.0
Obstetrician	16 (40)	9 (38)	7 (44)	
Certified nurse midwife	11 (28)	7 (29)	4 (25)	
Lay midwife	7 (18)	4 (17)	3 (19)	
Family practice doctor	4 (10)	3 (13)	1 (6)	
No prenatal care	2 (5)	1 (4)	1 (6)	
Perinatal complications†				
Hypertension	5 (12)	4 (16)	1 (6)	
Preterm labor	4 (10)	3 (12)	1 (6)	
Placental abruption	4 (10)	2 (8)	2 (12)	
Anemia	3 (7)	0	3 (19)	
Gestational diabetes	2 (5)	2 (8)	0	
Multiple pregnancy‡	2 (5)	2 (8)	0	
Postpartum infection	2 (5)	1 (4)	1 (6)	
Premature rupture of membranes	1 (2)	0	1 (6)	
Pyelonephritis	1 (2)	1 (4)	0	
Uterine rupture	1 (2)	1 (4)	0	
Substance use§				
Cigarettes	3 (7)	2 (8)	1 (6)	1.0
Alcohol	1 (2)	1 (4)	0	1.0
Recreational drugs	1 (2)	0	1 (6)	.6
Gestational age at delivery (wk±d)	38±6	37±9	39±5	.4
Location of birth				.6
Hospital	32 (78)	18 (72)	14 (88)	
Home	7 (17)	5 (20)	2 (13)	
Independent birth center	2 (5)	2 (8)	0	
Underwent labor induction	9 (22)	7 (28)	2 (12)	.3
Method of delivery				.5
Vaginal	29 (71)	16 (64)	13 (81)	
Cesarean	12 (30)	9 (36)	3 (19)	
Reason for cesarean delivery				.6
Previous cesarean delivery	1 (8)	1 (11)	0	
Breech presentation	1 (8)	1 (11)	0	
Placenta previa	1 (8)	1 (11)	0	
Arrest of labor	2 (17)	1 (11)	1 (33)	
Multiple pregnancy (twins)	1 (8)	1 (11)	0	
Requested cesarean delivery	3 (25)	3 (33)	0	
Other‖	3 (25)	1 (11)	2 (66)	
Birth weight (g)¶	3,146±1,671	2,914±1,276	3,490±625	.2
Neonate admitted to the NICU*	5 (14)	4 (20)	1 (7)	.4
Neonate diagnosed with an anomaly or developmental disorder*#	3 (9)	1 (5)	2 (14)	.7
Neonate diagnosed with a disorder of sexual development***	2 (6)	1 (5)	1 (7)	.8
Chest (breast) fed	21 (51)	10 (40)	11 (69)	.04

NICU, neonatal intensive care unit.
Data are n (%) or mean±standard deviation unless otherwise specified.
* Not all the participants answered this question.
† Includes complications occurring in the preconception, antepartum, intrapartum, and postpartum periods.
‡ Both sets of multiples were twins.
§ Survey question stated: "Once you knew you were pregnant, did you regularly: _ drink alcohol, _ smoke cigarettes, _ use recreational drugs, _ none of the above."
‖ Other reasons for cesarean delivery: placental abruption (n=1), preeclampsia (n=1), none specified (n=1).
¶ N=42 neonates resulting from a set of twins.
# Ventricular septal defect (n=1), bone cancer (n=1), sensory integration disorder (n=1).
** Intersex (n=1), micropenis (n=1).

noted that although their specific health care provider(s) may have been transgender-friendly, this was not necessarily the case with the office staff, nurses, and other health care workers.

DISCUSSION

The College has highlighted the need for obstetrician–gynecologists to help eliminate barriers to care for transgender men.[2] Our results demonstrate that transgender men desire children[4] and are willing and able to conceive, carry a pregnancy, and give birth. Participants repeatedly expressed a desire for more information regarding fertility options and access to reproductive health care providers who respect, support, and understand their gender identity.

Studies suggest that amenorrhea commonly occurs within 6 months of initiating testosterone therapy.[12,13] However, timeframe for resumption of menses after cessation of testosterone is unclear, and some have stated amenorrhea may be irreversible.[14] Participants who discontinued testosterone to attempt pregnancy reported resumption of menses within 6 months, with the majority within 3 months. Some conceived before return of menses. Despite small sample size, the timeline for menses resumption is consistent with that of literature on women who became amenorrheic with Sertoli-Leydig tumors and resumed menses after tumor resection.[15]

Although most transgender men in this study received prenatal care from a physician and delivered in a hospital, participants used nonphysician providers and nonhospital birth locations more frequently than the general public. In 2009, 99% of U.S. births occurred in hospitals,[16] compared with 78% of our participants. It is possible that health care provider choice and delivery location were responses to actual or anticipated negative experiences as suggested from many qualitative reports of suboptimal interactions with health care providers. However, health care provider and birth location may have resulted from other barriers such as access to health insurance.[17–20] Further research to clarify the experiences of transgender men with peripartum service provision will provide guidance for meeting their needs.

There is a 12% prevalence of major depressive disorders surrounding pregnancy, including postpartum depression, for women in the United States.[21] Although we did not specifically ask about depressive disorders, many of our participants reported experiences with peripartum depression in the narrative responses. A Canadian study of mental health among transgender men (n=207) found that depression was common.[22] Our findings suggest that transgender men may represent a high-risk population for postpartum depression and, although further research is warranted, future recommendations should emphasize assessment of peripartum depression in this population.

Nearly half of the transgender men who had not used testosterone had an unplanned pregnancy, a proportion comparable to that of the U.S. population.[23] Comparatively, one fourth of those previously on testosterone had unplanned pregnancies. By design this study cannot speak to incidence or prevalence of unplanned pregnancies among transgender men. However, given the financial burden[24] and risk of increased morbidity[25] from unintended pregnancy as well as the contraindication of testosterone use during pregnancy,[26,27] these findings suggest a potential unmet need for contraceptive services for transgender men.

Limitations to this study include those inherent with an online, cross-sectional survey, including not allowing for follow-up clarification from participants, decreasing responses from those with low literacy or other barriers to taking an online survey, and self-reported data raising concern for recall bias. The limited socioeconomic and racial diversity in respondents reduces immediate generalizability. Lastly, our eligibility criteria screened for transgender men who had a successful birth, impeding generalizable to those who attempt to get pregnant and cannot and those who do not carry to term. Strengths include the novelty of reporting transgender men's pregnancy experiences, inclusion of those who had socially and medically transitioned, and the mixed-methods format that allows insight into experiences.

Through demonstrating that transgender men are becoming pregnant and having babies, regardless of prior testosterone use, this preliminary study contributes data to emerging discussions regarding their reproductive health experiences. Respondents highlight the need for health care providers to partner with this community and develop gender-appropriate resources and support. Simple but meaningful steps for health care providers include establishing rapport by using patients' preferred names and pronouns, validating gender identity, and reflecting their individual relationships to their pregnancies. Counseling with transgender men should include discussions of reproductive goals, including fertility desires, and the role of contraception. We also suggest all health care providers discuss fertility preservation options with patients before initiating testosterone use in accordance with international standards of care.[26,27] More

clinical and investigational work is needed to understand the physical and emotional needs of transgender men during pregnancy and birth so that health care providers may partner with this underserved community to improve care. As we respond to calls for increased access to reproductive health care for transgender men, we must ensure that we can provide evidence-based, comprehensive services befitting their unique needs and concerns.[2]

REFERENCES

1. Grant JM, Mottet LA, Tanis J, Herman JL, Harrison J, Keisling M. National transgender discrimination survey report on health and healthcare. Washington (DC): National Center for Transgender Equality and National Gay and Lesbian Task Force; 2010. p. 1–23.
2. Healthcare for transgender individuals. Committee Opinion No. 512. American College of Obstetricians and Gynecologists. Obstet Gynecol 2011;118:1454–8.
3. Scheim A, Greta B. Sex and Gender Diversity Among Transgender Persons in Ontario, Canada: Results From a Respondent-Driven Sampling Survey. J Sex Res 2014;1–14 [Epub ahead of print].
4. Wierckx K, Van Caenegem E, Pennings G, Elaut E, Dedecker D, Van de Peer F, et al. Reproductive wish in transsexual men. Hum Reprod 2012;27:483–7.
5. Califia-Rice P. Family values: two dads with a difference –neither of us was born male. Village Voice June 20, 2000.
6. Hembree WC, Cohen-Kettenis P, Delemarre HA, Gooren LJ, Meyer WJ III, Spack NP, et al. Endocrine treatment of transsexual persons. Minneapolis (MN): WPATH (World Professional Association for Transgender Health); 2009.
7. Beatie T. Labor of love: the story of one mans extraordinary pregnancy. Berkeley (CA): Seal Press; 2008.
8. De Sutter P. Gender reassignment and assisted reproduction present and future reproductive options for transsexual people. Hum Reprod 2001;16:612–4.
9. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)-A metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform 2009;42:377–81.
10. Charmaz K. Constructing grounded theory: a practical guide through qualitative analysis. Thousand Oaks (CA): Pine Forge Press; 2006.
11. Pequegnat W, Rosser B, Bowen A, Bull S, DiClemente R, Bockting W, et al. Conducting internet-based HIV/STD prevention survey research: considerations in design and evaluation. AIDS Behav 2007;11:505–21.
12. Nakamura A, Watanabe M, Sugimoto M, Sako T, Mahmood S, Kaku H, et al. Dose-response analysis of testosterone replacement therapy in patients with female to male gender identity disorder. Endocr J 2013;60:275–81.
13. Steinle K. Hormonal management of the female-to-male transgender patient. J Midwifery Womens Health 2011;56:293–302.
14. T'Sjoen G, Van Caenegem E, Wierckx K. Transgenderism and reproduction. Curr Opin Endocrinol Diabetes Obes 2013;20:575–9.
15. Gui T, Cao D, Shen K, Yang J, Zhang Y, Yu Q, et al. A clinicopathological analysis of 40 cases of ovarian Sertoli–Leydig cell tumors. Gynecol Oncol 2012;127:384–9.
16. Martin JA, Hamilton BE, Ventura SJ, Osterman MJ, Kirmeyer S, Mathews TJ, et al. Births: final data for 2009. Natl Vital Stat Rep 2011;60:1–70.
17. Unger C. Care of the transgender patient: the role of the gynecologist. Am J Obstet Gynecol 2014;210:16–26.
18. Dutton L, Koenig K, Fennie K. Gynecologic care of the female-to-male transgender man. J Midwifery Womens Health 2008;53:331–7.
19. Poteat T, German D, Kerrigan D. Managing uncertainty: a grounded theory of stigma in transgender healthcare encounters. Soc Sci Med 2013;84:22–9.
20. Xavier J, Bradford J, Hendricks M, Safford L, McKee R, Martin E, et al. Transgender healthcare access in Virginia: a qualitative study. Int J Transgenderism 2013;14:3–17.
21. Le Strat Y, Dubertret C, Le Foll B. Prevalence and correlates of major depressive episode in pregnant and postpartum women in the United States. J Affect Disord 2011;135:128–38.
22. Rotondi NK, Bauer GR, Travers R, Travers A, Scanlon K, Kaay M. Depression in male-to-female transgender Ontarians: results from the trans PULSE project. Can J Commun Ment Health 2011;30:113–33.
23. Finer LB, Kost K. Unintended pregnancy rates at the state level. Perspect Sex Reprod Health 2011;43:78–87.
24. Trussell J, Henry N, Hassan F, Prezioso A, Law A, Filonenko A. Burden of unintended pregnancy in the United States: potential savings with increased use of long-acting reversible contraception. Contraception 2013;87:154–61.
25. Shah PS, Balkhair T, Ohlsson A, Beyene J, Scott F, Frick C. Intention to become pregnant and low birth weight and preterm birth: a systematic review. Matern Child Health J 2011;15:205–16.
26. Coleman E, Bockting W, Botzer M, Cohen-Kettenis P, DeCuypere G, Feldman J, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. Int J Transgenderism 2012;13:165–232.
27. Hembree WC, Cohen-Kettenis P, Delemarre-van de Waal HA, Gooren LJ, Meyer WJ III, Spack NP, et al. Endocrine treatment of transsexual persons: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2009;94:3132–54.
28. Kuper L, Nussbaum R, Mustanski B. Exploring the diversity of gender and sexual orientation identities in an online sample of transgender individuals. J Sex Res 2012;49:244–54.
29. Djordjevic M, Stanojevic D, Bizic M, Kojovic V, Majstorovic M, Vujovic S, et al. Metoidioplasty as a single stage sex reassignment surgery in female transsexuals: Belgrade experience. J Sex Med 2009;6:1306–13.

Assisted reproductive technology outcomes in female-to-male transgender patients compared with cisgender patients: a new frontier in reproductive medicine

Angela Leung, M.D.,[a,b] Denny Sakkas, Ph.D.,[a] Samuel Pang, M.D.,[a] Kim Thornton, M.D.,[a,b] and Nina Resetkova, M.D., M.B.A.[a,b]

[a] Boston IVF, Waltham; and [b] Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, Beth Israel Deaconess Medical Center, Boston, Massachusetts

Objective: To investigate assisted reproductive technology (ART) outcomes in a female-to-male transgender cohort and compare the results with those of a matched cisgender cohort.
Design: Matched retrospective cohort study.
Setting: In vitro fertilization clinic.
Patient(s): Female-to-male transgender patients (n = 26) who sought care from 2010 to 2018. A cisgender cohort (n = 130) was matched during the same time period by age, body mass index, and antimüllerian hormone levels.
Intervention(s): Not applicable.
Main Outcome Measure(s): Cycle outcomes, including oocyte yield, number of mature oocytes, total gonadotropin dose, and peak E_2 levels.
Result(s): The mean number of oocytes retrieved in the transgender group was 19.9 ± 8.7 compared with 15.9 ± 9.6 in the cisgender group. Peak E_2 levels were the same between the two groups. The total dose of gonadotropins used was higher in the transgender group compared with the cisgender group (3,892 IU vs. 2,599 IU). Of the 26 patients, 16 performed oocyte banking only. Seven couples had fresh or frozen transfers, with all achieving live births.
Conclusion(s): This is the first study of this size investigating ART outcomes in female-to-male transgender patients. The findings may serve to reassure transgender patients and their care providers that outcomes can be excellent even if testosterone therapy has already been initiated. Further investigation needs to be performed on the generalizability of these findings, and whether similar results can be achieved without stopping testosterone therapy. (Fertil Steril® 2019;112:858–65. ©2019 by American Society for Reproductive Medicine.)
El resumen está disponible en Español al final del artículo.

Key Words: Transgender, ovarian stimulation, fertility preservation

Transgender people represent a patient population that has previously received little attention, especially in the area of reproductive health and fertility. While the number of individuals who self-identify as transgender is small, surveys show that it is on the rise (1). In the U.S., 0.3%–2% of the population identify as transgender, although this may be grossly underestimated because tools to survey gender identity are poor (2, 3). The World Professional Association of Transgender Health (WPATH) estimates that worldwide, the prevalence for male-to-female individuals is 1:45,000 to 1:12,000, and for female-to-male individuals is 1:200,000 to 1:30,400 (4); however, most of the studies used to derive these numbers are from European countries.

Previously, there had been an assumption that transgender

Received May 10, 2019; revised June 21, 2019; accepted July 12, 2019; published online October 6, 2019.
A.L. has nothing to disclose. D.S. is on the Scientific Advisory Board of Cooper Surgical. S.P. has nothing to disclose. K.T. has nothing to disclose. N.R. has nothing to disclose.
Reprint requests: Angela Leung, M.D., Department of OB/GYN, Beth Israel Deaconess Medical Center, 330 Brookline Avenue, Kirstein 3, Boston, Massachusetts, 02215 (E-mail: aleung@bostonivf.com).

Fertility and Sterility® Vol. 112, No. 5, November 2019 0015-0282/$36.00

https://doi.org/10.1016/j.fertnstert.2019.07.014

individuals were not interested in maintaining their reproductive potential. There are also historical societal biases that the transgender population should not retain their reproductive potential. Until 2015, 24 countries in Europe required sterilization before changing gender assignment on legal documents (5, 6). That number has now decreased to 14 (7), even though the European Court of Human Rights ruled in 2017 that such requirements violate human rights law. These biases are changing, however; for example, after a Supreme Court ruling Germany's Cabinet recently approved a third gender option for official identification records. With this change in acceptance, the associated reproductive rights will also improve. Ethicists have concluded that there is no ethical basis to deny transgender individuals access to reproductive medicine (8–10). Both the American Society for Reproductive Medicine and the European Society of Human Reproduction and Embryology have issued opinions that transgender patients should have the same access to fertility options as cisgender patients and that fertility preservation options should be discussed before gender transition (9, 10).

Several recent studies have demonstrated that transgender people do desire parenthood, or at the least wish to preserve that possibility. A widely cited Belgian study in 2012 found that more than one-half (54%) of surveyed transgender men desired to have children, and 37% would have considered freezing oocytes if that option had been available (11). A German study found that 76% of both transgender men and women had thought about fertility preservation before transition, but only 9% of transgender women and 3% of transgender men had actually completed this process (12). Interestingly, more transgender women preferred to build families through adoption, whereas more transgender men desired to have biologic offspring and would prefer to build families through sexual intercourse or by carrying a pregnancy (13). Even studies in transgender teenagers have indicated that the desire for future family building exists in about one-half of them (14, 15). Therefore, it is safe to say that transgender individuals have similar interests in their reproductive potential as cisgender individuals.

This provides an exciting and much needed impetus to examine fertility options and outcomes for this marginalized patient population. The first and key intervention is the ability to preserve fertility through the cryopreservation of gametes before medical or surgical transition. In transgender men, this can be done via oocyte, embryo, or ovarian tissue cryopreservation (6, 16). In 2014, the first case report of oocyte cryopreservation in a transgender man was published (17). Since then, there has only been one other published case series of three transgender men who underwent fertility preservation and subsequent transfer of embryos after vitro fertilization (IVF) (18). There is a knowledge gap about how these patients respond to treatment and their experiences and outcomes with fertility preservation or IVF. As transgender individuals increasingly seek access to reproductive services, we seek to shed light on the optimal way to provide effective care to these patients.

Our objective in the present study was to examine cycle outcomes in a female-to-male transgender (also referred to as transgender male) cohort and compare the results with those of a matched cisgender cohort.

METHODS

Study Participants

We performed a database search of the electronic medical record eIVF (practicehwy.com) at a single large academically affiliated IVF clinic. Because our electronic medical record does not currently have the means to identify transgender individuals easily, we performed a text-based query. The query searched for any mention of the term "transgender" or other common derivatives of the term in the patients' charts. The query was conducted for January 2010 to July 2018, because the first transgender man that this clinic treated was in 2010. To be included in this study, the patient had to identify as a transgender man and have completed an ovarian stimulation cycle for oocyte cryopreservation, embryo cryopreservation, or intended uterine transfer. Most couples who desired to conceive did so through reciprocal IVF, whereby the transgender patient provided the oocytes and their cisgender partner carried the pregnancy. The few transgender men who opted to carry the pregnancy themselves underwent several failed intrauterine insemination cycles before proceeding to IVF. There were no exclusion criteria. Demographic data as well as cycle outcome data were collected.

All transgender patients who had already initiated androgen therapy stopped testosterone before cycle start and were instructed to wait for resumption of menses. Those patients who strongly opposed restarting menses instead had serum testosterone levels monitored until they returned to the upper levels of the normal female range. Baseline ovarian reserve parameters were checked once these criteria were met.

Ovarian stimulation protocol was determined by the treating physician. Almost all used an antagonist stimulation protocol, whereby dosing of gonadotropins was adjusted according to standard protocol in relation to the individual's response. These protocols were the same as those used for cisgender patients; there was no unique protocol specifically tailored for transgender patients.

Matching Procedure

In the comparison portion of our study, each transgender male patient was matched with five unique cisgender patients with either male-factor or tubal-factor infertility. This subset of patients was chosen for comparison in the assumption that they should have intact ovulatory function similar to that of transgender men who have not undergone transition. Cisgender patients with ovulatory dysfunction such as polycystic ovary syndrome (PCOS) were excluded. Only first ovarian stimulation cycles in both groups were used for comparison. Individual matching was performed with the use of age, body mass index (BMI), and antimüllerian hormone (AMH) levels. Age was matched by the Society for Assisted Reproductive Technology age categories (<30, 30 to <35, 35 to <38, 38 to <40, and ≥ 40 y), BMI by obesity class categories (25 to <30, 30 to <35, 35 to <40, and ≥ 40 kg/m^2), and AMH by clinically meaningful categories (<0.5, 0.5 to <1, and

≥1 ng/mL). Cycles were also matched according to next closest in cycle start date.

Data Analysis

The primary outcome in the matched cohort analysis was number of oocytes retrieved, and secondary outcomes included number of mature oocytes, total gonadotropin dose, and peak E_2 levels. Pregnancy outcomes are described in the transgender cohort, but not compared, because the small sample size of the transgender group would not produce statistically meaningful results.

Descriptive statistics were used to describe the transgender patient and cycle characteristics, as well as cycle outcomes. For the matching analysis, results were compared using Student *t* test, with *P* values $< .05$ defined as significant.

Ethical Approval

This study was approved by the Institutional Review Board at Beth Israel Deaconess Medical Center.

RESULTS

Characteristics of the Transgender Male Cohort

We identified a total of 53 transgender male patients who sought care from 2010 to 2018. Out of this group, 26 patients completed treatment; the remainder presented only for initial consultations. None of the patients who pursued treatment had undergone gender-affirming "bottom surgery" (i.e., oophorectomy or hysterectomy), so all patients included in this study had intact uteri and ovaries.

Twenty-nine cycles were completed by 26 patients, with patients completing a mean of 1.1 cycles. Seventeen cycles were for oocyte cryopreservation, five for IVF and embryo cryopreservation, and seven for IVF with embryo transfer.

The age range of patients was 14–39 years, with the average (±SD) age at cycle start being 28.3 ± 6.7. Those seeking oocyte cryopreservation were youngest, and those seeking embryo cryopreservation were oldest (Table 1A). All patients who created embryos had partners, whereas 50% of patients cryopreserving oocytes had partners. Only one patient already had a child through natural conception at the time of treatment. All other patients had never been pregnant.

A majority of patients (61%) were already undergoing testosterone therapy at the time of presentation, with the remainder planning transition after their assisted reproductive technology (ART) cycles. Types of testosterone therapy varied widely, including intramuscular injection, transdermal gel, and subcutaneous injection. Time on testosterone before seeking ART treatment ranged from 3 months to 17 years, with a mean of 44 months (3.7 years; Table 1A). All patients stopped testosterone before cycle start and almost all experienced resumption of menses. Baseline AMH, FSH, and E_2 levels and antral follicle counts (AFCs) were all normal in this patient group (Table 1A). On average, patients were off testosterone for 4 months before starting their cycle (range 1–12 months; Table 1A).

Transgender Cycle Outcomes

Across all of the transgender ovarian stimulation cycles (n = 29), a mean of 19.4 ± 8.4 oocytes were retrieved per cycle

TABLE 1

Characteristics of the transgender male cohort undergoing controlled ovarian hyperstimulation, by (A) patient and (B) cycle.

Patient variable	All patients (n = 26)	Oocyte cryopreservation (n = 16)	Embryo cryopreservation (n = 3)	IVF with transfer (n = 7)
Age (y)	28.3 ± 6.7	25.3 ± 6.2	35.6 ± 3.5	32.06 ± 4.2
BMI (kg/m^2)	26.0 ± 7.0	24.7 ± 7.3	29.4 ± 6.8	27.7 ± 6.4
AMH (ng/mL)	3.4 ± 1.9	3.6 ± 2.2	3.1 ± 1.4	2.7 ± 1.5
FSH (IU/mL)	6.2 ± 2.0	6.0 ± 2.0	6.2 ± 2.1	6.6 ± 2.5
AFC (n)	16.6 ± 5.3	17.2 ± 5.5	18.0 ± 3.6	14.7 ± 6.2
Has partner (%)	69.2	50	100	100
Initiated androgen therapy (%)	61.5	43.8	100	85.7
Time on testosterone (mo)	43.9 ± 31.0	39.7 ± 19.2	126 ± 110.3	48.0 ± 52.3
Time off testosterone (mo)	4.5 ± 3.5	4.4 ± 3.7	4.0 ± 2.6	5.0 ± 4.3
Resumed menses (%)	81.2	85.7	100	66.7
Time to cycle start (mo)	6.4 ± 12.8	2.9 ± 2.2	3.4 ± 3.2	15.6 ± 23.0
Cycle variable	**All cycles (n = 29)**	**Oocyte cryopreservation (n = 17)**	**Embryo cryopreservation (n = 5)**	**IVF with transfer (n = 7)**
Oocytes retrieved (n)	19.4 ± 8.4	22.7 ± 8.4	15.6 ± 7.3	14.3 ± 6.1
Mature oocytes (%)	75.4 ± 20.7	75.4 ± 22.5	59.0 ± 8.8	88.0 ± 11.7
Oocytes frozen (n)	–	17.7 ± 6.1	–	–
Embryos frozen (n)	4.2 ± 0.6	–	3.8 ± 4.2	4.7 ± 1.4
Embryos transferred (n)	–	–	–	1.3 ± 0.8
Pregnancy rate (%)[a,b]	–	–	–	83.3
Live birth rate (%)[b]	–	–	–	58.3

Note: Data are presented as mean ± standard deviation or %. AFC = antral follicle count; AMH = antimüllerian hormone; BMI = body mass index; IVF = in vitro fertilization.
[a] Clinical pregnancy defined as positive fetal heart beat on ultrasound.
[b] Rates calculated per transfer, including fresh and frozen (see Supplemental Table 1 [available online at www.fertstert.org] for more details).

Leung. ART outcomes in transgender male patients. Fertil Steril 2019.

(Table 1B). In the 17 oocyte cryopreservation cycles, a mean of 17.7 ± 6.1 oocytes were cryopreserved per cycle. A mean of 3.8 ± 4.2 embryos were cryopreserved in the five embryo cryopreservation cycles (Table 1B).

Among the patients who planned for IVF with transfer, two intended to carry the pregnancy themselves and the remaining five transferred embryos to their cisgender female partner. A total of 12 transfers were performed among the seven patient couples: five fresh and seven frozen-thawed embryo transfers (FETs). The pregnancy and live birth rates per transfer are summarized in Table 1B, and detailed cycle outcomes are listed in Supplemental Table 1 (available online at www.fertstert.org). Of the five patient pairs who completed a fresh transfer, four became pregnant and delivered. Two patients who had live births from their fresh transfers subsequently returned a few years later for FET. These resulted in current ongoing pregnancies. One patient had a miscarriage following fresh embryo transfer, went on to have a FET, and ultimately conceived and had a live birth. Two patients underwent cryopreserve-all cycles for preimplantation genetic testing for aneuploidy and subsequent FET. Both patients conceived and delivered from their FET cycles. Single-embryo transfer was performed in nine of 12 cycles; two embryos were transferred in three of 12 cycles. A mean of 4.7 ± 1.4 supernumerary embryos were frozen (Table 1B). All seven patient pairs who transferred embryos ultimately had successful outcomes with pregnancy and delivery of healthy children: six singletons and one set of twins.

Matched Analysis

First cycles only of the transgender cohort (n = 26) were then matched with 130 cisgender cycles (Table 2). Of these cisgender cycles, 80% had a diagnosis of male-factor infertility and 20% tubal-factor infertility. Age, BMI, and AMH were well matched across the two groups, with no significant difference in any of these categories (Table 2).

On average, more oocytes were retrieved in the transgender cycles compared with cisgender cycles, and this result was statistically significant (19.9 ± 8.7 vs. 15.9 ± 9.6; P=.04; Table 2). Number of mature oocytes and peak E_2 levels were similar between the two groups; however, significantly higher total doses of gonadotropins were used in the transgender stimulation cycles (Table 2).

A subanalysis was performed on only the transgender patients who had initiated testosterone therapy (Table 3), which showed similar results. The exception was that although the number of oocytes retrieved from these transgender patients also trended higher, the difference was not statistically significant.

DISCUSSION

It has been posited that transgender men must undergo fertility preservation before any hormonal therapy to achieve good results (16). However, our study shows that these patients can have ovarian stimulation outcomes that are similar to those of cisgender counterparts, and this seems to be true even in cases of patients who have already initiated hormonal transition with the use of testosterone.

This is the first study to describe transgender cycle parameters and outcomes in such detail and scope. The only other publication that describes outcomes in this population is a case series of three transgender men who underwent fertility preservation before gender-affirming hormonal therapy: one patient cryopreserved oocytes only, and two cryopreserved oocytes and returned for IVF and embryo transfer (18). More than 18 oocytes were retrieved in each of those cases, and both patients who underwent embryo transfer (to their cisgender female partner) conceived and delivered. Our findings support the preliminary positive outcomes seen in that study. The transgender patients in our cohort had a mean of 20 oocytes retrieved, and all who transferred embryos eventually achieved a successful pregnancy and delivery.

Compared with the matched cisgender patients, the transgender male patients performed well. The mean number of oocytes retrieved was higher in the transgender group than in the cisgender group. It should be kept in mind that the primary purpose of this study was to explore overall ART outcomes, and given the retrospective nature of the study and limited sample size, it was not powered to compare egg yield in transgender and cisgender patients. Therefore, the significant difference we found should be interpreted conservatively. However, the trend of good oocyte yield in the transgender cohort is very reassuring for this patient population.

Although more than one-half the transgender patients had been on testosterone therapy before undergoing the

TABLE 2

Comparison of cisgender and transgender ovarian stimulation cycles.

Variable	Cisgender (n = 130)	Transgender (n = 26)	*P* value
Age (y)	30.4 ± 3.8	28.3 ± 6.7	.12
AMH (ng/mL)	4.0 ± 3.7	3.4 ± 1.9	.22
BMI (kg/m^2)	26.2 ± 6.6	26.0 ± 7.0	.90
Oocytes retrieved (n)	15.9 ± 9.6	19.9 ± 8.7	.04
Mature oocytes (%)	82.1 ± 17.1	78.3 ± 20.3	.38
Peak E_2 level (pg/mL)	2,715.9 ± 1515.2	2,755.5 ± 1,297.7	.89
Total gonadotropin dose (IU)	2,599.1 ± 1,313.6	3,891.8 ± 1,577.6	< .001

Note: Data presented as mean ± standard deviation. Abbreviations as in Table 1.

Leung. ART outcomes in transgender male patients. Fertil Steril 2019.

TABLE 3

Comparison of transgender patients with previous androgen exposure and matched cisgender cycles.

Variable	Cisgender (n = 80)	TG with androgen exposure (n = 16)	*P* value
Oocytes retrieved (n)	14.4 ± 8.9	18.6 ± 9.3	.11
Mature oocytes (%)	84.4 ± 16.1	77.0 ± 23.3	.24
Peak E_2 level (pg/mL)	2,713.2 ± 1,487.4	2,943.1 ± 1,364.7	.55
Total gonadotropin dose (IU)	2,707.0 ± 1,452.1	4,155.5 ± 1,507.6	.002

Note: Data presented as mean ± standard deviation.

Leung. ART outcomes in transgender male patients. Fertil Steril 2019.

stimulation cycles (the longest duration being 17 years), all of them had discontinued testosterone for a short period immediately before cycling. It may be hypothesized that the subset of transgender patients who had testosterone therapy would have poorer egg yields compared with matched cisgender patients, but our subanalysis found this not to be true. These transgender patients still had excellent stimulation outcomes, with an average of 18 oocytes retrieved. These results suggest that even long periods of gender-affirming androgen therapy do not appear to have negative effects on ovarian stimulation outcomes. These clinical findings echo experiments that show that androgen treatment does not reduce the ovarian follicle pool or cortical distribution (19–21).

One possible explanation for why transgender patients may produce more oocytes is that their biochemical environment is like that of a woman with PCOS owing to their higher levels of circulating androgens. It has been shown that androgen excess can accelerate the growth of early follicular development while simultaneously slowing the rate of atresia of early antral follicles (22). This in turn can lead to the polycystic ovary morphology. Experiments in mice (23) and nonhuman primates (24) that involve exogenous injection of androgens support this theory. Consequently, patients with PCOS typically exhibit evidence of high ovarian reserve, with high values of AMH and AFC. They have a robust response to ovarian stimulation, and a high number of oocytes are usually retrieved.

However, this theory was not borne out in a study by Ikeda et al. (19), who described that excessive androgen exposure in transgender men did not lead to a polycystic ovary morphology. Therefore, this theory cannot fully explain the high oocyte yield in the transgender group. Unfortunately, testosterone levels were not measured in all of the transgender patients in our study. In the few who were tested after discontinuing androgen therapy, their levels were within upper parameters of the normal female range before cycle start. In fact, to start their cycle, the patient must have either resumed menses or have had a testosterone level in the normal female range. At this time, we do not fully understand the repercussions that long-term androgen therapy may have on the ovarian environment, and this is certainly an area of needed study. But at least in the time immediately before the stimulation cycle, there were no clinical signs that the transgender patients had higher than normal levels of androgens. Furthermore, their ovarian reserve testing was similar to that of our cisgender group and did not exhibit the higher levels of AMH and AFC that are typical of PCOS patients.

Our data also showed that significantly higher total doses of gonadotropins were used in the stimulation cycles of transgender patients. An underlying reason for this difference may be because for most of these transgender patients, an oocyte cryopreservation cycle is a "one-shot deal." In some cases, insurance will pay for only one cycle, or if self-paying they can only afford only one cycle; for others, they had mentally braced themselves to stop testosterone therapy only long enough to undergo one cycle. Therefore, in many cases aggressive stimulation was intentional to optimize egg yield for a single cycle. Consequently, it is unclear whether transgender male patients truly require more medication than their cisgender counterparts. It may be argued that the transgender group's high oocyte yield is partly due to aggressive stimulation, and lower doses may eliminate the difference we found between the two groups. However, data from examining individual cycles in detail seem to indicate that transgender patients still have good egg yields with low doses of medication. In our study, the two transgender cycles that used the lowest total dose of gonadotropins (<1,750 IU) both had more than 25 oocytes retrieved. In contrast, the cisgender cycles that also used a total dose of <1,750 IU (n = 37) had a mean of 16 oocytes retrieved. The sample size here is too small to draw definitive conclusions, but the tendency does suggest that aggressive stimulation alone cannot explain the difference we found in oocyte yield.

Age is also an important factor that must be considered. It is well established that younger patients have higher oocyte yields (25–27). The population of transgender patients undergoing ovarian stimulation is generally younger than that of the infertility population. In our cohort, the youngest patient was 14 years old and had been referred for fertility preservation before starting androgen therapy. The number of younger patients in our study reflects a robust referral system that is sensitive to the fertility desires of transgender patients. This is an important aspect of transgender medicine that improves access, but the younger average age of this population makes a comparative study difficult because most do not carry an infertility diagnosis. Ultimately, a better comparison group would be oncology-fertility patients, because many of these patients present at younger ages at the time of fertility preservation, similarly to our transgender group. They may, however, have other

underlying issues owing to their diagnosis. A second, but slightly less optimal, comparison group would be cisgender female patients undergoing elective oocyte cryopreservation. However, even this group might be slightly older than the transgender cohort, because teenagers are not offered elective oocyte cryopreservation. At the time of our study, we did not have a sufficient sample size of oncology-fertility or elective oocyte cryopreservation patients to perform an adequately matched analysis. Regardless, we attempted to mitigate the effect of age through our matching process, and the resulting comparison showed that age was not statistically different in the two groups. Therefore, our finding of excellent oocyte yields in the transgender group can not be explained solely by the influence of age.

Another limitation of our data is the selection of patients with infertility as our comparison group. We used tubal-factor and male-factor infertility patients only, excluding all patients with ovulatory dysfunction, with the presumption that this subset of patients closely mimics the transgender cohort where personal choices preclude egg and sperm meeting in a natural environment. However, this presumption may not necessarily be true, because transgender patients are not truly infertile, especially those who have not initiated hormone therapy. On the other hand, even though the cisgender patients may have a documented case of tubal-factor or male-factor infertility, there may be other factors contributing to their infertility that were not revealed in the initial workup. The presence of these unknown confounders may skew the cisgender group to a slightly worse outcome and could contribute to the difference in oocyte yield that was found between the two groups.

Although this study population is the largest cohort of transgender male patients to be described in the literature thus far, it is interesting to note that it represents only about 50% of the transgender male patients who presented for fertility consultation at our clinic. Many of the patients who ultimately did not choose to proceed with treatment did so because of the need to stop testosterone therapy before initiating a cycle or the burden of cost. For many transgender patients, stopping androgen therapy can be both physically and psychologically distressing, especially because many experience the resumption of menses. This concern about halting or even delaying start of hormone therapy is often cited as a reason to defer fertility preservation (14, 28, 29). It is important to note, however, that some patients in our cohort were offered the option of deferring menses start before treatment, and the resulting outcomes were in line with the remainder of the cohort. A logical follow-up question is whether ovarian stimulation can be done with any measure of success without the cessation of testosterone. Although our findings are certainly reassuring for patients who have already initiated androgens, they were still all required to stop therapy to proceed with stimulation. This is a barrier to access that should be investigated, and if overcome may increase utilization of ART by transgender male patients. Given the retrospective nature of this study, we were also not able to assess the experiences of these patients who presented for treatment. In the future, concomitant quantitative and qualitative investigation may help to clarify transgender patients' perspectives on treatment and identify additional barriers to care.

CONCLUSION

Female-to-male transgender patients who choose to access ART for fertility preservation or family building can have excellent results. Patients who have already started the transition process via hormone therapy should be reassured that they still have the opportunity to preserve fertility as long as they retain their ovaries. Providers can counsel these patients that oocyte yield is on par with that of their cisgender counterparts. And although outcome data from patients who transferred an embryo is limited, preliminary findings suggest a high rate of success.

Acknowledgments: The authors thank Michele Hacker, Sc.D., M.S.P.H., at Beth Israel Deaconess Medical Center for providing statistical consultation; and coauthor Samuel Pang, M.D., at Boston IVF for being the managing physician for the majority of the patients included in this study.

REFERENCES

1. Meerwijk EL, Sevelius JM. Transgender population size in the united states: a meta-regression of population-based probability samples. Am J Public Health 2017;107:1–8.
2. Gates G. How many people are lesbian, gay, bisexual and transgender? Williams Institute; 2011. Available at: https://williamsinstitute.law.ucla.edu/research/census-lgbt-demographics-studies/how-many-people-are-lesbian-gay-bisexual-and-transgender/. Accessed September 13, 2018.
3. Flores AR, Herman JL, Gates GJ, Brown TNT. How many adults identify as transgender in the United States. Williams Institute; 2016. Available at: https://williamsinstitute.law.ucla.edu/research/how-many-adults-identify-as-transgender-in-the-united-states/. Accessed September 13, 2018.
4. Hunger S. Commentary: Transgender people are not that different after all. Camb Q Healthc Ethics 2012;21:287–9.
5. Nixon L. The right to (trans) parent: a reproductive justice approach to reproductive rights, fertility, and family-building issues facing transgender people. William Mary J Race Gend Soc Justice 2013;20:73.
6. t'Sjoen G, van Caenegem E, Wierckx K. Transgenderism and reproduction. Curr Opin Endocrinol Diabetes Obes 2013;20:575–9.
7. Transgender Europe. Trans rights Europe map 2018: forced sterilization. Available at: https://tgeu.org/wp-content/uploads/2018/05/MapB_TGEU2018_Online.pdf.
8. Murphy TF. The ethics of fertility preservation in transgender body modifications. J Bioethical Inq 2012;9:311–6.
9. Ethics Committee of the American Society for Reproductive Medicine. Access to fertility services by transgender persons: an Ethics Committee opinion. Fertil Steril 2015;104:1111–5.
10. de Wert G, Dondorp W, Shenfield F, Barri P, Devroey P, Diedrich K, et al. ESHRE Task Force on Ethics and Law 23: medically assisted reproduction in singles, lesbian and gay couples, and transsexual people. Hum Reprod 2014;29:1859–65.
11. Wierckx K, van Caenegem E, Pennings G, Elaut E, Dedecker D, van de Peer F, et al. Reproductive wish in transsexual men. Hum Reprod 2012;27:483–7.
12. Auer MK, Fuss J, Nieder TO, Briken P, Biedermann SV, Stalla GK, et al. Desire to have children among transgender people in Germany: A cross-sectional multi-center study. J Sex Med 2018;15:757–67.
13. Tornello SL, Bos H. Parenting intentions among transgender individuals. LGBT Health 2017;4:115–20.
14. Nahata L, Tishelman AC, Caltabellotta NM, Quinn GP. Low fertility preservation utilization among transgender youth. J Adolesc Health 2017;61:40–4.
15. Strang JF, Jarin J, Call D, Clark B, Wallace GL, Anthony LG, et al. Transgender youth fertility attitudes questionnaire: measure development in nonautistic

and autistic transgender youth and their parents. J Adolesc Health 2018;62:128–35.
16. de Roo C, Tilleman K, T'Sjoen G, de Sutter P. Fertility options in transgender people. Int Rev Psychiatry 2016;28:112–9.
17. Wallace SA, Blough KL, Kondapalli LA. Fertility preservation in the transgender patient: expanding oncofertility care beyond cancer. Gynecol Endocrinol 2014;30:868–71.
18. Maxwell S, Noyes N, Keefe D, Berkeley AS, Goldman KN. Pregnancy outcomes after fertility preservation in transgender men. Obstet Gynecol 2017;129:1031–4.
19. Ikeda K, Baba T, Noguchi H, Nagasawa K, Endo T, Kiya T, et al. Excessive androgen exposure in female-to-male transsexual persons of reproductive age induces hyperplasia of the ovarian cortex and stroma but not polycystic ovary morphology. Hum Reprod 2013;28:453–61.
20. van den Broecke R, van der Elst J, Liu J, Hovatta O, Dhont M. The female-to-male transsexual patient: a source of human ovarian cortical tissue for experimental use. Hum Reprod 2001;16:145–7.
21. de Roo C, Lierman S, Tilleman K, Peynshaert K, Braeckmans K, Caanen M, et al. Ovarian tissue cryopreservation in female-to-male transgender people: insights into ovarian histology and physiology after prolonged androgen treatment. Reprod Biomed Online 2017;34:557–66.
22. Homburg R. Androgen circle of polycystic ovary syndrome. Hum Reprod 2009;24:1548–55.
23. Beloosesky R, Gold R, Almog B, Sasson R, Dantes A, Land-Bracha A, et al. Induction of polycystic ovary by testosterone in immature female rats: modulation of apoptosis and attenuation of glucose/insulin ratio. Int J Mol Med 2004;14:207–15.
24. Abbott DH, Barnett DK, Bruns CM, Dumesic DA. Androgen excess fetal programming of female reproduction: a developmental aetiology for polycystic ovary syndrome? Hum Reprod Update 2005;11:357–74.
25. American College of Obstetricians and Gynecologists, American Society for Reproductive Medicine. Female age-related fertility decline. Committee opinion no. 589. Fertil Steril 2014;101:633–4.
26. Chuang C-C, Chen C-D, Chao K-H, Chen S-U, Ho H-N, Yang Y-S. Age is a better predictor of pregnancy potential than basal follicle-stimulating hormone levels in women undergoing in vitro fertilization. Fertil Steril 2003;79:63–8.
27. Hull MGR, Fleming CF, Hughes AO, McDermott A. The age-related decline in female fecundity: a quantitative controlled study of implanting capacity and survival of individual embryos after in vitro fertilization. Fertil Steril 1996;65:783–90.
28. Armuand G, Dhejne C, Olofsson JI, Rodriguez-Wallberg KA. Transgender men's experiences of fertility preservation: a qualitative study. Hum Reprod 2017;32:383–90.
29. Chen D, Simons L, Johnson EK, Lockart BA, Finlayson C. Fertility preservation for transgender adolescents. J Adolesc Health 2017;61:120–3.

Resultados de las Técnicas de Reproducción Asistida en pacientes transgénero mujer-hombre comparados con los de pacientes cisgénero: una nueva frontera en Medicina Reproductiva

Objetivo: Investigar los resultados de las Técnicas de Reproducción Asistida (ART) en una cohorte transgénero mujer-hombre y comparar los resultados con los de una cohorte emparejada de pacientes cisgénero

Diseño: Estudio cruzado retrospectivo de cohortes.

Entorno: Clínica de Fecundación in Vitro.

Paciente(s): Pacientes transgénero mujer-hombre (n=26) que solicitaron tratamiento desde 2010 a 2018. Una cohorte de pacientes cisgénero (n=130) fue emparajeda por edad, índice de masa corporal y niveles de hormona antimulleriana durante el mismo periodo de tiempo.

Intervención(es): No aplica.

Medida del Resultado Principal: Resultados del ciclo, incluyendo número total de ovocitos, número de ovocitos maduros, dosis total de gonadotrofinas y pico máximo de estradiol.

Resultado(s): El número medio de ovocitos recuperados en el grupo transgénero fue de 19.9 +/- 8.7 comparado con 15.9 +/-9.6 en el grupo cisgénero. Los niveles pico de estradiol fueron similares entre los dos grupos. La dosis total de gonadotrofinas utilizadas fue mayor en el grupo transgénero comparado con el grupo cisgénero (3,892 IU vs. 2,599 IU). De los 36 pacientes, 16 realizaron únicamente criopreservación de ovocitos. Siete parejas realizaron transferencias en fresco o de embriones congelados, consiguiendo todas recién nacidos vivos.

Conclusión(es): Este es el primer estudio con este número de pacientes investigando los resultados de ART en pacientes transgénero mujer-hombre. Los resultados pueden servir para confirmar a los pacientes transgénero y a sus médicos que los resultados pueden ser excelentes incluso si el tratamiento con testosterona ya ha sido iniciado. Es necesario realizar más estudios sobre la posible generalización de estos hallazgos y para ver si resultados similares pueden obtenerse sin cesar el tratamiento con testosterona.

Transgender Health
Volume 5, Number 3, 2020

DOI: 10.1089/trgh.2019.0077

SHORT REPORT

Changes in Anxiety and Depression from Intake to First Follow-Up Among Transgender Youth in a Pediatric Endocrinology Clinic

Annette L. Cantu,[1,*] Danielle N. Moyer,[1] Kara J. Connelly,[2] and Amy L. Holley[1]

Abstract

Monitoring acute distress in transgender youth initiating gender-affirming care is important given their increased risk for significant mental health symptoms. The current study examined changes in anxiety, depression, and suicidality from initial appointment to first follow-up in 80 youth, ages 11–18. Average time between visits was ~4 months but varied across participants. Results revealed no change in acute distress from intake to follow-up. Neither distance from medical center nor initiation of hormone therapy was associated with symptom changes. While research shows decreased distress with initiation of hormones, study findings suggest changes may actually take longer to occur.

Keywords: transgender, acute distress, mental health, behavioral health screeners, gender dysphoria, access to care

Introduction

The Patient Health Questionnaire-9 (PHQ-9) for depression and anxiety (Generalized Anxiety Disorder-7 [GAD-7]) are brief, easy-to-use, physician-administered screening measures used to identify acute distress among transgender and gender-nonconforming (TGN) youth.[1] Given TGN youth are at higher risk for anxiety, depression, and suicidal ideation than their peers,[2] identifying youth who endorse high levels of acute distress in the gender clinic setting can highlight those who need access to mental health services and crisis interventions. In our recent study, examining rates of depression, anxiety, and suicidal ideation in TGN youth, 43% of patients 11–18 years of age endorsed clinically significant depression symptoms, 61% of patients endorsed clinically significant anxiety symptoms, and 30% of patients endorsed thoughts of death or self-harm on several days or greater.[1]

Medical interventions for pubertal adolescents fall into three general categories: (1) medications to suppress or manage the estrogen or testosterone produced by the body (e.g., hormone blockers), (2) hormone therapy (HT) to masculinize or feminize the body, and (3) gender-affirming surgeries.[3] Some research suggests a positive impact of HT on the mental health of TGN youth and relief of gender dysphoria over time with initiation of surgical interventions.[2,4] A gap in knowledge exists in how symptoms of acute distress change in the short-term, particularly in youth receiving gender-affirming care who are not undergoing surgery. Gathering these data on how mood changes early in treatment is important for informing both patients and providers on what to expect.

Another gap in the literature is how distance from a medical center may be associated with mood symptomatology. Mental health differences have been found among TGN adults who live in rural versus nonrural areas.[5] TGN youth and adults also consistently report access to TGN specialists, including mental health providers, as a common barrier to health care.[6–8] While there are 53 comprehensive clinical care programs for TGN youth in the United States and Canada, the majority are located in large metropolitan cities, making access to services challenging for many families.[9] To our knowledge, no research has explored whether distance to a comprehensive clinical care program for

Divisions of [1]Psychology and [2]Endocrinology, Department of Pediatrics, Oregon Health and Science University, Portland, Oregon, USA.

*Address correspondence to: Annette L. Cantu, PhD, Division of Psychology, Department of Pediatrics, Oregon Health and Science University, 707 SW Gaines Street, Portland, OR 97239, USA, E-mail: cantua@ohsu.edu

TGN youth might impact the level and persistence of acute distress among these youth.

To fill these gaps in the research, the current study has three aims: (1) Describe changes in anxiety, depression, and suicidality from intake visit to first follow-up appointment using the PHQ-9 and GAD-7 in TGN youth; (2) Examine symptoms of anxiety, depression, and suicidality by distance from medical center (>30 vs. <30 miles) from intake to follow-up; (3) Examine changes in anxiety and depression from intake to first follow-up among the TGN youth who initiate HT.

Methods

Participants and procedure

Participants were TGN youth seeking gender-affirming care at an academic medical center in the Northwestern United States between September 2017 and June 2019. All youth ages 11 and older complete anxiety and depression screeners at every visit regardless of mental health diagnoses or symptom severity. In this clinic, youth do not receive prescriptions for hormone medication management at the initial visit, but many patients initiate medications between their first and second appointments after completing required steps (family receives extensive counseling, signs consent form, completes assessment, and acquires a letter of support from an experienced mental health provider). Initiation of hormone management medications (e.g., hormone blockers) and HT is individualized for each patient and some patients initiate both at the same time. Second visit is recommended 3–4 months after the initial visit.

Youth were included in the current study if they (1) were between the ages of 11 and 18 years, (2) had attended both an initial visit and one follow-up appointment, and (3) completed measures assessing acute distress (PHQ-9 and GAD-7) at both visits. Retrospective chart review was used to extract patient age, affirmed gender, medical interventions, screener results, and distance from clinic. Because chart review was used to collect data, no informed consent procedures were conducted, and data on participant race/ethnicity, socioeconomic status, and education level were not available to include in analyses. It was also infeasible to document the exact time of HT initiation given the variability in how and where individuals received their treatments. The institution's Human Subjects Institutional Review Board approved all study procedures.

Measures

Depression. The PHQ-9[10] is a 9-item screening measure of depression. Items are rated on a 4-point scale for how often each symptom has occurred in the past 2 weeks from 0 (Not at All) to 3 (Nearly Every Day). The final item (item 9) asks about thoughts of death and self-harm. Youth were coded as endorsing suicidal ideation if they responded ≥1 on item 9.

Anxiety. The GAD-7[11] is a 7-item screening measure of anxiety. Items are rated on a 4-point scale for how often each symptom has occurred in the past 2 weeks from 0 (Not at All) to 3 (Nearly Every Day).

Distance from the medical center. Distance from the participants' home to the medical center was calculated using the zip code of the youth's home address documented in the medical record. This variable was then dichotomized into two groups: youth who lived <30 miles and youth who lived >30 miles to the medical center.

Analyses

Descriptive statistics were used to examine sample characteristics and differences in endorsements of suicidal ideation. Paired sample *t*-tests were used to examine overall changes from initial visit to follow-up, and independent sample *t*-tests were used to examine simple differences across groups. Repeated measures factorial analysis of variance (ANOVA) was used to examine the role of potential moderators (i.e., initiation of HT and distance from clinic) in the changes in distress over time. All analyses were completed using SPSS, with the exception of power analyses, which were completed using G*Power.

Results

Subject demographics are depicted in Table 1. Participants were 80 youth 11–18 years of age (80 youth completed PHQ-9 screeners at both time points and 78 youth completed GAD-7 screeners at both time points). Average time between initial visit and follow-up appointment was 4.7 months. However, there was significant variability ranging from <1 month to 11 months, with 80% of follow-up visits occurring between 2 and 7 months.

Only 1 individual initiated HT before the initial visit, and 28 youth initiated HT between initial visit and first follow-up. Of those 28 youth, 6 were started on feminizing hormones and 22 were started on masculinizing

Table 1. Summary of Sample Characteristics ($n = 80$)

Age in years, mean (SD), range	15.1 (1.8)
Distance in miles, mean (SD), range	36.2 (39.9)
Within 30 miles, *n* (%)	46 (57.5)
Beyond 30 miles, *n* (%)	34 (42.5)
Affirmed gender, *n* (%)	
Female	15 (18.8)
Male	58 (72.5)
Nonbinary	7 (8.8)
Follow-up time in weeks, mean (SD)	20.4 (10.2)

Interventions, *n* (%)	Initial visit	Follow-up
Hormone blockers only	2 (2.5)	13 (16.2)
HT only	1 (1.3)	25 (31.2)
Both hormone blockers and HT	0 (0.0)	4 (5.0)
Neither hormone blockers or HT	77 (96.2)	38 (47.5)

Interventions represent treatment initiated before current visit.
HT, hormone therapy.

Table 2. Change in Depression and Anxiety from Initial Visit to First Follow-Up

	Initial visit Mean (SD)	Follow-up Mean (SD)	*t*	*p*
Total sample				
PHQ-9 ($n = 80$)	10.5 (6.5)	10.0 (6.4)	0.87	0.385
GAD-7 ($n = 78$)	9.1 (6.1)	8.8 (5.7)	0.58	0.561
	Initial visit Mean (SD)	Follow-up Mean (SD)	*F*	*p*
PHQ-9 distance			1.33	0.253
Within 30 miles ($n = 46$)	10.5 (6.9)	10.6 (6.7)		
Beyond 30 miles ($n = 34$)	10.5 (6.2)	9.3 (6.1)		
GAD-7 distance			2.44	0.123
Within 30 miles ($n = 44$)	8.7 (6.3)	9.1 (5.6)		
Beyond 30 miles ($n = 34$)	9.6 (5.8)	8.4 (5.9)		
PHQ-9 HT			1.44	0.235
HT initiated ($n = 28$)	9.8 (7.1)	10.3 (7.3)		
No HT ($n = 51$)	11.1 (6.3)	10.1 (5.9)		
GAD-7 HT			0.24	0.624
HT initiated ($n = 27$)	8.4 (6.4)	8.5 (5.5)		
No HT ($n = 50$)	9.6 (5.9)	9.1 (5.8)		

GAD-7, Generalized Anxiety Disorder-7; HT, hormone therapy; PHQ-9, Patient Health Questionnaire-9.

hormones. A total of 17 youth initiated hormone blockers between their initial visit and first follow-up, 4 of which also initiated HT. No analyses were conducted to examine differences between which hormone was started or for those that also initiated puberty blockers due to small and skewed samples.

Aim 1 examined changes in anxiety, depression, and suicidality from initial visit to first follow-up appointment. At initial visit, 37 (46%) youth met the cutoff for depression (PHQ-9 score ≥ 11). Of those 37 youth, 28 (76%) continued to meet the cutoff at follow-up and 9 (24%) no longer met the cutoff for clinically significant depression. At initial visit, 49 (61%) youth met the cutoff for anxiety (GAD-7 ≥ 6). Of those 49 youth, 41 (84%) continued to meet the cutoff at follow-up and 8 (16%) no longer met the cutoff for clinically significant anxiety. For the total sample, mean values for anxiety and depression were lower at first follow-up compared with initial appointment, but these changes were not statistically significant (Table 2). Changes in suicidality from initial visit to follow-up were also examined. Of the 27 (34%) youth who endorsed suicidality at intake, 22 (81%) continued to endorse suicidality at their follow-up visit, and only 4 (4%) no longer endorsed suicidality at follow-up.

Aim 2 examined changes in acute distress symptoms for those living within 30 miles of the medical center compared with those living beyond 30 miles (Table 2). A repeated measures factorial ANOVA did not reveal any significant differences in anxiety or depression by proximity to the clinic. Mean changes were examined qualitatively for potential trends, and a similar pattern of interaction effects was observed across depression and anxiety scores. Specifically, small decreases in mean scores on anxiety and depression were observed from initial visit to follow-up, but only for the youth living beyond 30 miles. There were no differences in the number of youth who endorsed suicidal ideation at intake or follow-up visit by proximity from clinic.

Aim 3 examined changes in acute distress symptoms for those who initiated HT between initial visit and first follow-up compared with those who did not begin treatment (Table 2). Participants started on hormone blockers only were not included in this analysis. The analysis *included* the four youth who were started on both HT and hormone blockers and *excluded* the participant who had started HT before the initial visit. A repeated measures factorial ANOVA did not reveal any significant differences in depression and anxiety scores among youth who did versus did not initiate HT following their intake visit. Similarly, there were no differences in the endorsement of suicidal ideation between initial and follow-up visit for youth who did versus did not initiate gender-affirming hormones.

Conclusion

The present study explored acute changes in depression, anxiety, and suicidal ideation from initial visit in a pediatric gender clinic to first follow-up visit among TGN youth, with attention to differences

between those living within and beyond 30 miles of the clinic and between those who did versus did not initiate HT between visits. Depression and anxiety were assessed using the PHQ-9 and GAD-7, which are completed by all youth 11 years and older at every clinic visit. These screening measures were chosen due to their ease of use, sensitivity to change, and scoring that can be used clinically and for research.[10] Previously published data from this clinic has demonstrated that these screeners capture rates of anxiety and depression similar to the broader literature.[1]

Overall, the results of this study suggest that no clinically significant changes in mood symptoms occur during this initial time frame, with the majority of youth maintaining similar levels of symptomatology at their first follow-up as they did at their initial visit. While some evidence to date lends strong support for symptom improvement over time[4,12–16] the current study suggests changes likely occur gradually and may not begin to occur until several months into treatment. This may be related to the fact that masculinizing and feminizing physical changes occur slowly after initiation of HT.[17] While most previous research has used more in-depth psychological assessments and interviews rather than screeners, the PHQ-9 and GAD-7 are sensitive to change[10] and have been shown in this clinic to capture similar rates of anxiety and depression as broader literature,[1] lending further support to the strength of these results.

These results lend support for educating providers, youth, and their families about setting appropriate expectations for change. The results of this study suggest that this overall pattern of symptom maintenance in the early stages of treatment did not differ between those who initiated HT before their first follow-up and those who did not. This suggests that improvements in mood symptoms with HT may take longer to occur or that other factors not assessed in the current study (e.g., level of family support, access to mental health services) play a more significant role in early improvements.

The present study also examined the role of proximity to clinic in changes in distress. There were no significant differences in distress at either time point, or changes over time by distance suggesting that this particular measurement of clinic access did not impact early changes in distress. With larger samples it will be important to look at differences in youth who live in urban versus rural areas and also use more robust measures to assess access to clinical care and resources. It will be important in future studies to assess distance as a potential proxy for community support as well as access to services.

There are important limitations to take into account when interpreting these results. Data collected were limited to one clinic, with a relatively small sample size and only two time points examined. Power analyses revealed that the current sample would have been well powered to detect large effects, but not small-to-moderate effects, which are more likely when looking at shorter time frames. Due to sample size, we could not examine how age, affirmed gender, or initiation of hormone blockers were associated with changes in symptoms of distress. Sample size also limits the ability to examine differences among those with genderqueer and nonbinary identities.

Another important limitation includes the variability with regard to exactly when HT was initiated and the variability in time between initial visit and first follow-up. The recommended follow-up time between first and second visits is typically 3–4 months, and most participants in this study actually attended a second appointment 2 to 7 months after the initial visit. While variability may impact the results of this study, these data reflect what is typical for this clinic, which likely generalizes to other real-world situations. Given that patients may not experience the physical effects of HT until 3–6 months following initiation,[17] depression and anxiety may not be impacted until visible effects begin to occur.

Further research is required to determine the trajectory of mood during the course of HT over multiple time points as well as how symptoms change with variable factors (e.g., mental health treatment, etc.). Data continues to be collected within this clinic and analyses with larger samples and additional time points are planned which will help increase our understanding of mental health changes during treatment in this population.

Author Disclosure Statement

No competing financial interests exist.

Funding Information

No funding was received for this article.

References

1. Moyer D, Connelly K, Holley A. Using the PHQ-9 and GAD-7 to screen for acute distress in transgender youth: findings from a pediatric endocrinology clinic. J Pediatr Endocrinol Metab. 2016;32:71–74.
2. Connolly M, Zervos M, Barone C, et al. The mental health of transgender youth: advances in understanding. J Adolesc Health. 2016;59:489–495.

3. Coleman E, Bockting W, Botzer M, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. Int J Transgend. 2012;13:165–232.
4. De Vries A, McGuire J, Steensma T, et al. Young adult psychological outcome after puberty suppression and gender reassignment. Pediatrics. 2014;134:696–704.
5. Horvath K, Iantaffi A, Swinburne-Romine R, Bockting W. A comparison of mental health, substance use, and sexual risk behaviors between rural and non-rural transgender persons. J Homosex. 2014;61:1117–1130.
6. Sanchez F, Sanchez J, Danoff A. Health care utilization, barriers to care, and hormone usage among male-to-female transgender persons in New York City. Am J Public Health. 2009;99:713–719.
7. Rider G, McMorris B, Gower A, et al. Health and care utilization of transgender and gender nonconforming youth: a population-based study. Pediatrics. 2018;141:e20171683.
8. Safer J, Coleman E, Feldman J, et al. Barriers to health care for transgender individuals. Curr Opin Endocrinol Diabetes Obes. 2016;23:168–171.
9. Human Rights Campaign. Interactive map: clinical care programs for gender-expansive children and adolescents. www.hrc.org/resources/interactive-map-clinical-care-programs-for-gender-nonconforming-childr, Accessed May 2019.
10. Richardson L, McCauley E, Grossman D, et al. Evaluation of the Patient Health Questionnaire (PHQ-9) for detecting major depression among adolescents. Pediatrics. 2010;126:1117–1123.
11. Löwe B, Decker O, Müller S, et al. Validation and standardization of the Generalized Anxiety Disorder Screener (GAD-7) in the general population. Med Care. 2008;46:266–274.
12. Colizzi M, Costa R, Todarello O. Transsexual patients' psychiatric comorbidity and positive effect of cross-sex hormonal treatment on mental health: results from a longitudinal study. Psychoneuroendocrinology. 2014;39:65–73.
13. Gorin-Lazard A, Baumstarck K, Boyer L, et al. Hormonal therapy is associated with better self-esteem, mood, and quality of life in transsexuals. J Nerv Ment Dis. 2013;201:996–1000.
14. Manieri C, Castellano E, Crespi C, et al. Medical treatment of subjects with gender identity disorder: the experience in an Italian public health center. Int J Transgend. 2014;15:53–65.
15. Nguyen H, Chavez A, Lipner E, et al. Gender-affirming hormone use in transgender individuals: impact on behavioral health and cognition. Curr Psychiatry Rep. 2018;20:110.
16. Tucker R, Testa R, Simpson T, et al. Hormone therapy, gender affirmation surgery, and their association with recent suicidal ideation and depression symptoms in transgender veterans. Psychol Med. 2018;48:2329–2336.
17. Hembree WC, Cohen-Kettenis PT, Gooren L, et al. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an endocrine society clinical practice guideline. J Clin Endocrinol Metab. 2017;102:3869–3903.

Cite this article as: Cantu AL, Moyer DN, Connelly KJ, Holley AL (2020) Changes in anxiety and depression from intake to first follow-up among transgender youth in a pediatric endocrinology clinic, *Transgender Health* 5:3, 196–200, DOI: 10.1089/trgh.2019.0077.

Abbreviations Used

ANOVA = analysis of variance
GAD-7 = Generalized Anxiety Disorder-7
HT = hormone therapy
PHQ-9 = Patient Health Questionnaire-9
SD = standard deviation
TGN = transgender and gender nonconforming

Fertility preservation in transgender men without discontinuation of testosterone

Brett A. Stark, M.D., M.P.H. and Evelyn Mok-Lin, M.D.

Division of Reproductive Endocrinology and Infertility, Department of Obstetrics, Gynecology and Reproductive Sciences, University of California San Francisco; San Francisco, California.

Objective: To report two cases of fertility preservation in two transgender men without an extended period of higher dose testosterone cessation.
Design: Chart abstraction was completed for two cases of oocyte preservation in transgender men without stopping testosterone gender-affirming therapy before controlled ovarian stimulation (COS).
Setting: A university-affiliated fertility clinic in San Francisco, California.
Patient(s): Two 27-year-old transgender men on higher dose testosterone undergoing oocyte cryopreservation.
Intervention(s): Not applicable.
Main Outcome Measure(s): Both patients had been on 6 and 20 months of testosterone therapy, respectively, and continued throughout COS. A random start antagonist plus letrozole protocol was used for the patient in case 1, with a leuprolide acetate trigger. A luteal start antagonist protocol was applied to the patient in case 2 with a leuprolide acetate trigger.
Result(s): In case 1, a total of 35 oocytes were retrieved, with a total of 23 metaphase II (MII) oocytes cryopreserved. An additional 7 MII oocytes were obtained after in vitro maturation for a total of 30 MII oocytes that were vitrified. In case 2, 14 oocytes were retrieved, and 9 mature oocytes (MII) were vitrified.
Conclusion(s): Transgender men have historically been advised to discontinue testosterone before COS, a process that may be distressing for many individuals. This is the first published case report demonstrating the proof of concept of COS without cessation of high-dose testosterone therapy in two transgender men. Future studies with larger sample sizes should be performed to confirm these findings. (Fertil Steril Rep® 2022;3:153–6. ©2022 by American Society for Reproductive Medicine.)
Key Words: Fertility preservation, transgender, ovarian stimulation, testosterone, oocyte preservation

The World Professional Association for Transgender Health, Endocrine Society, and American Society of Reproduction all recommend counseling transgender men on Assisted Reproductive Technologies (ART) and fertility preservation (FP) before initiation of gender-affirming treatment (GAT) (1). For postpubertal transgender men who present after initiating GAT with testosterone therapy, data on best practices for FP are limited because our current understanding of the long-term impact of testosterone therapy on reproduction is poorly understood and largely speculative.

The current practice, due to lack of data on controlled ovarian stimulation (COS) and oocyte outcomes while continuing high-dose testosterone therapy, is to temporarily suspend testosterone treatment for an arbitrary length of time, usually between 1 and 6 months or until the resumption of menses (2–4). However, COS involves significant "female" hormone exposure with associated physical symptoms, frequent monitoring with transvaginal ultrasound, and transvaginal aspiration of oocytes under sedation. While these procedures alone can be traumatic to some transgender men, the physical changes associated with discontinuation of testosterone and female hormonal stimulation can be significantly dysphoric and a possible barrier to those seeking FP (4, 5).

One published case report demonstrated successful COS with leuprolide acetate injection for final maturation of the oocytes in a 20-year-old transgender man who had been on testosterone therapy for 18 months (6). At the time of retrieval, the patient was on 25 mg of weekly intramuscular testosterone, and a total of 22

Received September 17, 2021; revised and accepted February 3, 2022.
B.A.S. has nothing to disclose. E.M.-L. has nothing to disclose.
Reprint requests: Brett A. Stark, M.D., M.P.H., Department of Obstetrics, Gynecology & Reproductive Sciences, University of California, San Francisco, 490 Illinois Street, 10th Floor, Box 0132, SF, California 94158 (E-mail: brett.stark@ucsf.edu).

Fertil Steril Rep® Vol. 3, No. 2, June 2022 2666-3341

https://doi.org/10.1016/j.xfre.2022.02.002

metaphase II oocytes were cryopreserved (6). According to standard guidelines for the management of masculinizing hormone therapy, typical dosing of intramuscular/subcutaneous testosterone cypionate for transgender men is 50 mg/week, with a maximum dosage of 100 mg/week (7). Lower dosages of testosterone, starting as low as 20 mg/week intramuscularly/subcutaneously, are recommended for genderqueer and nonbinary individuals (7).

In our clinic, patients presenting on testosterone therapy at the time that they decide to proceed with COS are informed of the unknown effects of testosterone on the ability of the ovary to respond to gonadotropin stimulation, oocyte quality, the ability of these oocytes to fertilize, live birth rates, and potential long-term epigenetic effects on offspring. Patients are recommended to withhold testosterone treatment for 1–3 months before initiation of COS. Despite this counseling, patients still may elect to proceed with COS without cessation of testosterone therapy.

We report herein two cases of oocyte preservation in transgender men who elected to undergo COS for FP without cessation of GAT with higher dosages of testosterone cypionate therapy day.

CASE REPORT

With each patient's written informed consent, we conducted a retrospective chart review of two patients who underwent ART for FP without cessation of GAT with testosterone from 2020 to 2021. All data was obtained from chart review and reported without any patient identifiers. Institutional review board exemption from our institution was obtained for this study.

Case 1

A 27-year-old transgender man who had been taking weekly testosterone injections since April of 2019 was referred for FP counseling in May of 2020 in preparation for a gender-affirming hysterectomy and bilateral salpingo-oophorectomy. He was nulliparous and had been amenorrheic since May of 2019. Before starting testosterone, he had regular 28-day cycles. He endorsed a remote history of oral contraceptive pills used for birth control. His medical history was notable for diabetes and hypertension. He had undergone bilateral mastectomy. He was taking subcutaneous testosterone cypionate 80 mg weekly. On examination, his vital signs were normal, and his body mass index was 25.06 kg/m^2. A transvaginal ultrasound showed an anteverted uterus (volume, 33 cm^3), an endometrial stripe of 5.3 mm with normal ovaries bilaterally, and an antral follicle count (AFC) of 36. His serum testosterone level at the time of presentation was 1,273 ng/dL, and his serum antimüllerian hormone level was 11.9 ng/mL. The options for FP were reviewed in detail, and the patient expressed interest in oocyte cryopreservation.

He decided to proceed with FP in January 2021 and continued testosterone (80 mg subcutaneously) throughout the process. Given the presence of amenorrhea, a random start protocol was initiated with subcutaneous follitropin alfa (150 IU; Gonal-F, Merck Canada) and subcutaneous menotropins (150 IU; Menopur, Ferring Canada). Letrozole (5 mg orally; Femara) was given daily throughout the stimulation to maintain low estradiol (E2) levels for the purpose of minimizing the potential dysphoria associated with elevated levels and potential withdrawal bleeding on completion of the cycle. On stimulation day 4, his dosage of menotropins was decreased to 75 IU subcutaneously. Daily subcutaneous ganirelix acetate (0.25 mg; Orgalutran, Merck) was initiated on stimulation day 8 until the day of trigger. Because of a robust response, his follitropin alfa (Gonal-F, Merck Canada) dose was decreased to 75 IU subcutaneously on stimulation day 9. At this time, the patient's E2 was noted to be 1,381 pg/mL, and he endorses symptoms of abdominal bloating. Given his desire to maintain physiologically low levels of estrogen, letrozole was increased to 7.5 mg orally on stimulation day 11.

Follicle tracking was performed by transvaginal ultrasound without difficulty. When the lead follicle reached 20 mm, with most follicles in the 13–20 mm range, a subcutaneous leuprolide acetate trigger (4 mg [0.8 mL]) was given. Laboratory values before the trigger included luteinizing hormone (LH) levels of 5.97 IU/L and E2 levels of 1371 pg/mL. Post-trigger laboratory findings revealed an appropriate response to the agonist trigger with LH levels of 67.23 IU/L and progesterone levels (p4) of 12.3 nmol/L. The endometrium achieved a thickness of 8.6 mm. There were a total of 35 oocytes retrieved, with a total of 23 MII oocytes cryopreserved on the day of retrieval. An additional 7 oocytes progressed to MII 1-day postretrieval with in vitro maturation for a total of 30 MII oocytes that were vitrified. The patient reported no major side effects related to the ovarian stimulation aside from mild abdominal cramping and bloating.

Case 2

A 27-year-old transgender man who has been taking weekly testosterone injections since August of 2020 presented in September 2020 for FP counseling. He was nulliparous with regular 28-day menses even before the initiation of testosterone therapy. He denied previous hormonal contraceptive pill use. He was healthy and was preparing to undergo a bilateral mastectomy. He was taking testosterone cypionate (60 mg subcutaneously weekly) with plans to increase his dose on completion of ART (to 80 mg subcutaneously weekly). On examination, his vital signs were normal, and his body mass index was 29.05 kg/m^2. A transvaginal ultrasound showed an anteverted uterus (volume, 35 cm^3), an endometrial stripe of 3.8 mm with normal ovaries bilaterally, and an AFC of 9. His serum T level at the time of presentation was 410 ng/dL, and his antimüllerian hormone level was 2.67 ng/mL. The options for FP were reviewed in detail, and the patient expressed interest in oocyte cryopreservation.

In February of 2021, with the continuation of his weekly testosterone, he started follitropin alfa (300 IU; Gonal-F, Merck Canada) and menotropins (150 IU; Menopur, Ferring Canada) subcutaneously after completion of a baseline ultrasound and confirmation of entrance to the luteal phase. The patient was counseled on the use of letrozole (5 mg orally; Femara) to maintain low E2 levels for the purpose of minimizing the potential dysphoria associated, but he declined. Daily

ganirelix acetate (0.25 mg; Orgalutran, Merck) was injected starting on stimulation day 5 until the day of final oocyte maturation. Follicle tracking was performed by transvaginal ultrasound without difficulty. When the lead follicle reached 22 mm with the majority in the 13–18 mm range on stimulation day 9, a leuprolide acetate subcutaneous trigger (4 mg [0.8 mL]) was given. On the day of trigger, his max E2 level was 1,749 pg/mL and LH level was 2.11 IU/L. One day after the trigger, his laboratory values included an LH level of 19.09 IU/L and a progesterone level of 3.3 nmol/L. Given the low-normal levels in response to agonist trigger, a chorionic gonadotropin (5,000 IU subcutaneously; Pregnyl, Merck Canada) booster was given that evening and he proceeded with an oocyte retrieval 36 hours after the agonist trigger. The endometrium achieved a thickness of 8.8 mm. There were a total of 14 oocytes retrieved, and 9 MII oocytes were vitrified. The remainder of the oocytes were germinal vesicles, and postretrieval in vitro maturation was unsuccessful. The patient tolerated the process well and reported no major side effects of ovarian stimulation aside from mild abdominal bloating.

DISCUSSION

This is the first case report demonstrating the proof of concept of COS for FP in transgender men without cessation of typical to high-dose testosterone therapy. In our current case report, the patients in the described cases were on 6–20 months of testosterone before undergoing oocyte cryopreservation. The dosages described in these two cases are higher than the level observed in the previously described case study (6).

Many parallels can be made to FP for transgender men and oncofertility patients. To increase the chance for success in oncofertility, COS is typically performed with high doses of gonadotropins to maximize the number of oocytes retrieved and stored (8). In our clinic, a similar approach is often used with transgender men in an attempt to reduce the burden and potential gender dysphoria associated with multiple rounds of COS. Interestingly, studies evaluating ovarian histological changes after testosterone exposure in birth-assigned females have reported an ovarian phenotype similar to polycystic ovary syndrome–polycystic follicles with increased AFC with increased collagenization of the tunica albuginea, stromal hyperplasia, and luteinization of stromal cells (9–11). Patients with this ovarian morphology, particularly with a high AFC, as seen in patient 1 of our series, are known to be at higher risk for ovarian hyperstimulation syndrome (12). To balance the desire of maximizing success with as few COS cycles as possible and the risk of OHSS, we routinely implement antagonist protocols with leuprolide acetate trigger to reduce the risk of OHSS in this theoretically high-risk patient population (12). Prior research has shown that in TM populations with testosterone exposure, antagonist-based protocols are a feasible means of ovarian stimulation (13).

It is notable that the patient in case 1, who was receiving higher doses of testosterone at the time of COS and had evidence of higher systemic testosterone levels, had a particularly robust response to agonist trigger in comparison to the patient in case 2 of our study. In the previously described case report of one transgender male undergoing COS without cessation of lower dose testosterone, the authors noted a blunted response to agonist trigger and brought into question the ability of the pituitary to mount a physiologic response after prolonged testosterone exposure (6).

As COS is associated with exposure to supraphysiological levels of estrogen, a significant concern exists regarding the safety of the procedure in patients with hormone-sensitive cancers (14, 15). As such, the use of letrozole in conjunction with classic COS protocols has been advocated to avoid unnecessary and potentially harmful effects associated with the rise in estrogen levels on cancer (16, 17). The COS with letrozole was associated with significantly decreased peak estradiol levels without any negative impact on the number of mature oocytes collected (18). We use a similar approach with transmasculine individuals in our clinic, routinely counseling patients on the potential benefits of letrozole. While letrozole does, to an extent, limit our ability to track follicular growth, it decreases the individual's exposure to estrogen and the potentially dysphoria-inducing symptoms, including posttreatment withdrawal bleeding. The patient in case 2 opted to not proceed with letrozole therapy as he had recently started testosterone therapy and was not yet amenorrheic. Prior studies have shown there is a dose-dependent amenorrheic response to testosterone and, while >90% of transmasculine people on testosterone achieve amenorrhea by 6 months, menses can persist for up to a year or longer (19, 20).

While COS has historically been a viable option for many transgender men, it is not without major limitations. Little is known regarding the long-term impact of testosterone exposure on embryo quality, fertilization, pregnancy outcomes, and long-term outcomes from offspring. A study by Lierman et al. (21) from 2017 assessed the developmental competence of testosterone-exposed oocytes in transgender men. In this study of 16 transgender men, the authors found that the spindle structure analysis, a qualitative marker for oocyte functionality, and chromosomal alignment after vitrification appeared normal (21).

To the author's knowledge, no relevant animal studies or case reports of pregnancies using androgen-exposed oocytes without cessation of testosterone during COS have been described, and our current understanding of the long-term impact of testosterone exposure on reproductive outcomes is largely speculative. In a cross-sectional study of 41 transgender men who became pregnant and delivered after transition, 5 transgender men became unintentionally pregnant while amenorrhoeic on testosterone (3). While the detailed length of time on testosterone was not described for these 5 individuals, data from this study as a whole argues that transgender men on testosterone can retain fertility and become pregnant (3).

Two recent studies report outcomes of transgender men with a history of testosterone use after temporary discontinuation of testosterone before COS. Adeleye et al. (13) reported on COS outcomes in a cohort of 13 transgender men, 7 with a history of testosterone use for a median of 46 months. Notably, 3 transgender men with prior testosterone use presented for further family planning, with 2 desiring transfer

of embryos with donor insemination into cisgender female partners and 1 desiring autologous transfer of embryos inseminated with cisgender male partner's sperm. All 3 couples became successfully pregnant. Leung et al. (22) reported on ART outcomes in 26 transgender men, 61% of whom had been on testosterone from 3 months to 17 years. Seven couples desiring pregnancy were described; all 7 ultimately became pregnant with deliveries of healthy children. While small and retrospective in nature, both of these studies suggest that follicular development and oocyte quality do not seem to be significantly impacted by prior testosterone use (13, 22).

CONCLUSION

We present two cases of transgender men undergoing COS without cessation of testosterone GAT. Both patients in the reported cases had adequate responses to COS while continuing 60–80 mg of testosterone therapy. Additionally, our patient on 20 months of testosterone had a robust response to an agonist-only trigger. This case report adds to the small body of literature exploring the necessity of stopping testosterone therapy before the initiation of ART in transgender men. Continuation of testosterone may improve the experience of transgender men and decrease gender dysphoria exacerbation that has previously been described with COS. Additional outcomes, including fertilization rates, embryo quality, pregnancy and live birth rates, and long-term outcomes for offspring, should be further investigated.

REFERENCES

1. Coleman E, Bockting W, Botzer M, Cohen-Kettenis P, DeCuypere G, Feldman J, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, Version 7. Int J Transgend 2012;13:165–232.
2. De Roo C, Tilleman K, T'Sjoen G, De Sutter P. Fertility options in transgender people. Int Rev Psychiatry 2016;28:112–9.
3. Light AD, Obedin-Maliver J, Sevelius JM, Kerns JL. Transgender men who experienced pregnancy after female-to-male gender transitioning. Obstet Gynecol 2014;124:1120–7.
4. Armuand G, Dhejne C, Olofsson JI, Rodriguez-Wallberg KA. Transgender men's experiences of fertility preservation: a qualitative study. Hum Reprod 2017;32:383–90.
5. Wierckx K, Elaut E, Van Hoorde B, Heylens G, De Cuypere G, Monstrey S, et al. Sexual desire in trans persons: associations with sex reassignment treatment. J Sex Med 2014;11:107–18.
6. Gale J, Magee B, Forsyth-Greig A, Visram H, Jackson A. Oocyte cryopreservation in a transgender man on long-term testosterone therapy: a case report. F S Rep 2021;2:249–51.
7. Deutsch MB. Guidelines for the primary and gender-affirming care of transgender and gender nonbinary people. UCSF Transgender Care, Department of Family and Community Medicine, University of California San Francisco; 2016. Available at: https://transcare.ucsf.edu/guidelines. Accessed September 9, 2021.
8. Sunkara SK, Rittenberg V, Raine-Fenning N, Bhattacharya S, Zamora J, Coomarasamy A. Association between the number of eggs and live birth in IVF treatment: an analysis of 400 135 treatment cycles. Hum Reprod 2011;26:1768–74.
9. Ikeda K, Baba T, Noguchi H, Nagasawa K, Endo T, Kiya T, et al. Excessive androgen exposure in female-to-male transsexual persons of reproductive age induces hyperplasia of the ovarian cortex and stroma but not polycystic ovary morphology. Hum Reprod 2013;28:453–61.
10. Futterweit W, Deligdisch L. Histopathological effects of exogenously administered testosterone in 19 female to male transsexuals. J Clin Endocrinol Metab 1986;62:16–21.
11. Chadha S, Pache TD, Huikeshoven FJM, Brinkmann AO, van derKwast TH. Androgen receptor expression in human ovarian and uterine tissue of long-term androgen-treated transsexual women. Hum Pathol 1994;25:1198–204.
12. Mourad S, Brown J, Farquhar C. Interventions for the prevention of OHSS in ART cycles: an overview of Cochrane reviews. Cochrane Database Syst Rev 2017.
13. Adeleye AJ, Cedars MI, Smith J, Mok-Lin E. Ovarian stimulation for fertility preservation or family building in a cohort of transgender men. J Assist Reprod Genet 2019;36:2155–61.
14. Lambertini M, Pescio MC, Viglietti G, Goldrat O, Mastro LD, Anserini P, et al. Methods of controlled ovarian stimulation for embryo/oocyte cryopreservation in breast cancer patients. Expert Rev Qual Life Cancer Care 2017;2:47–9.
15. Lambertini M, Di Maio M, Pagani O, Curigliano G, Poggio F, Del Mastro L, et al. The BCY3/BCC 2017 survey on physicians' knowledge, attitudes and practice towards fertility and pregnancy-related issues in young breast cancer patients. Breast 2018;42:41–9.
16. Oktay K, Buyuk E, Libertella N, Akar M, Rosenwaks Z. Fertility preservation in breast cancer patients: a prospective controlled comparison of ovarian stimulation with tamoxifen and letrozole for embryo cryopreservation. J Clin Oncol 2005;23:4347–53.
17. Meirow D, Raanani H, Maman E, Paluch-Shimon S, Shapira M, Cohen Y, et al. Tamoxifen co-administration during controlled ovarian hyperstimulation for in vitro fertilization in breast cancer patients increases the safety of fertility-preservation treatment strategies. Fertil Steril 2014;102:488–95.e3.
18. Bonardi B, Massarotti C, Bruzzone M, Goldrat O, Mangili G, Anserini P, et al. Efficacy and safety of controlled ovarian stimulation with or without letrozole co-administration for fertility preservation: a systematic review and meta-analysis. Front Oncol 2020;10:574669.
19. Nakamura A, Watanabe M, Sugimoto M, Sako T, Mahmood S, Kaku H, et al. Dose-response analysis of testosterone replacement therapy in patients with female to male gender identity disorder. Endocr J 2013;60:275–81.
20. Spratt DI, Stewart I, Savage C, Craig W, Spack NP, Chandler DW, et al. Subcutaneous injection of testosterone is an effective and preferred alternative to intramuscular injection: demonstration in female-to-male transgender patients. J Clin Endocrinol Metab 2017;102:2349–55.
21. Lierman S, Tilleman K, Braeckmans K, Peynshaert K, Weyers S, T'Sjoen G, et al. Fertility preservation for trans men: frozen-thawed in vitro matured oocytes collected at the time of ovarian tissue processing exhibit normal meiotic spindles. J Assist Reprod Genet 2017;34:1449–56.
22. Leung A, Sakkas D, Pang S, Thornton K, Resetkova N. Assisted reproductive technology outcomes in female-to-male transgender patients compared with cisgender patients: a new frontier in reproductive medicine. Fertil Steril 2019;112:858–65.

Cochrane Library

Cochrane Database of Systematic Reviews

Antiandrogen or estradiol treatment or both during hormone therapy in transitioning transgender women (Review)

Haupt C, Henke M, Kutschmar A, Hauser B, Baldinger S, Saenz SR, Schreiber G

Haupt C, Henke M, Kutschmar A, Hauser B, Baldinger S, Saenz SR, Schreiber G.
Antiandrogen or estradiol treatment or both during hormone therapy in transitioning transgender women.
Cochrane Database of Systematic Reviews 2020, Issue 11. Art. No.: CD013138.
DOI: 10.1002/14651858.CD013138.pub2.

www.cochranelibrary.com

TABLE OF CONTENTS

Cochrane Library
Trusted evidence.
Informed decisions.
Better health.

Cochrane Database of Systematic Reviews

[Intervention Review]

Antiandrogen or estradiol treatment or both during hormone therapy in transitioning transgender women

Claudia Haupt[1], Miriam Henke[2], Alexia Kutschmar[3], Birgit Hauser[4], Sandra Baldinger[5], Sarah Rafaela Saenz[6], Gerhard Schreiber[7]

[1]Medical Service for Transgenders Lucerne, Lucerne, Switzerland. [2]Independent Transgender Consultant, Rothenburg, Germany. [3]Independent Transgender Consultant, Gaimersheim, Germany. [4]Gynaecology and Obstetrics, HRT Transgender Center Medical Practice, Hohenstein, Germany. [5]Independent Transgender Consultant, Lucerne, Switzerland. [6]Independent Transgender Consultant, Ludwigshafen, Germany. [7]Institute for Theology and Social Ethics, Technical University Darmstadt, Darmstadt, Germany

Contact address: Claudia Haupt, ddr.haupt@hin.ch.

Editorial group: Cochrane Tobacco Addiction Group.
Publication status and date: New, published in Issue 11, 2020.

Citation: Haupt C, Henke M, Kutschmar A, Hauser B, Baldinger S, Saenz SR, Schreiber G. Antiandrogen or estradiol treatment or both during hormone therapy in transitioning transgender women. *Cochrane Database of Systematic Reviews* 2020, Issue 11. Art. No.: CD013138. DOI: 10.1002/14651858.CD013138.pub2.

ABSTRACT

Background

Gender dysphoria is described as a mismatch between an individual's experienced or expressed gender and their assigned gender, based on primary or secondary sexual characteristics. Gender dysphoria can be associated with clinically significant psychological distress and may result in a desire to change sexual characteristics. The process of adapting a person's sexual characteristics to their desired sex is called ‘transition.'

Current guidelines suggest hormonal and, if needed, surgical intervention to aid transition in transgender women, i.e. persons who aim to transition from male to female. In adults, hormone therapy aims to reverse the body's male attributes and to support the development of female attributes. It usually includes estradiol, antiandrogens, or a combination of both. Many individuals first receive hormone therapy alone, without surgical interventions. However, this is not always sufficient to change such attributes as facial bone structure, breasts, and genitalia, as desired. For these transgender women, surgery may then be used to support transition.

Objectives

We aimed to assess the efficacy and safety of hormone therapy with antiandrogens, estradiol, or both, compared to each other or placebo, in transgender women in transition.

Search methods

We searched MEDLINE, the Cochrane Central Register of Controlled Trials (CENTRAL), Embase, Biosis Preview, PsycINFO, and PSYNDEX. We carried out our final searches on 19 December 2019.

Selection criteria

We aimed to include randomised controlled trials (RCTs), quasi-RCTs, and cohort studies that enrolled transgender women, age 16 years and over, in transition from male to female. Eligible studies investigated antiandrogen and estradiol hormone therapies alone or in combination, in comparison to another form of the active intervention, or placebo control.

Data collection and analysis

We used standard methodological procedures expected by Cochrane to establish study eligibility.

Main results

Our database searches identified 1057 references, and after removing duplicates we screened 787 of these. We checked 13 studies for eligibility at the full text screening stage. We excluded 12 studies and identified one as an ongoing study. We did not identify any completed studies that met our inclusion criteria. The single ongoing study is an RCT conducted in Thailand, comparing estradiol valerate plus cyproterone treatment with estradiol valerate plus spironolactone treatment. The primary outcome will be testosterone level at three month follow-up.

Authors' conclusions

We found insufficient evidence to determine the efficacy or safety of hormonal treatment approaches for transgender women in transition. This lack of studies shows a gap between current clinical practice and clinical research. Robust RCTs and controlled cohort studies are needed to assess the benefits and harms of hormone therapy (used alone or in combination) for transgender women in transition. Studies should specifically focus on short-, medium-, and long-term adverse effects, quality of life, and participant satisfaction with the change in male to female body characteristics of antiandrogen and estradiol therapy alone, and in combination. They should also focus on the relative effects of these hormones when administered orally, transdermally, and intramuscularly. We will include non-controlled cohort studies in the next iteration of this review, as our review has shown that such studies provide the highest quality evidence currently available in the field. We will take into account methodological limitations when doing so.

PLAIN LANGUAGE SUMMARY

Does hormone therapy help transgender women undergoing gender reassignment to transition?

Background

Transgender women may feel that they have been born in a body with the wrong sexual characteristics. This may result in significant psychological distress (gender dysphoria) and the desire to adapt their male physical and sexual characteristics to be more consistent with their experienced female gender. This is a process called transition. If measures to aid transition are not taken, this can result in greater psychological distress. One of the medical treatments given to help transgender women with male bodies to achieve transition is synthetic female hormones. These hormones can be taken by mouth, absorbed through the skin or injected into muscle.

Study characteristics

We looked for randomised controlled trials (RCTs) that included transgender women (age 16 and over) in transition from male to female. RCTs are a type of research study that can reduce the possibility of several types of bias. To be included in this review, studies needed to compare different hormone treatments used to support transgender women to transition (oestrogen alone, testosterone blockers alone, or oestrogen in combination with testosterone blockers), or compare these hormone treatments to placebos (fake or dummy treatments that appear to be the same as the actual treatment, but have no medical effects). We wanted to see whether hormone treatments help transgender women to make a transition that they are happy with. We also wanted to look at whether there were any health risks of the treatment.

Key results

We searched for studies up to 19 December 2019. We were unable to find any relevant completed studies that we could include. We did find one ongoing study that aimed to recruit all of the people taking part in the study by the end of 2020. This study is comparing the effects of estradiol valerate plus cyproterone treatment with estradiol valerate plus spironolactone treatment in transitioning transgender women in Thailand.

Quality of evidence

Our review found no RCTs that looked at whether hormone therapies are effective and safe when used to help transgender women to transition. Therefore, high-quality RCTs are needed to research these questions.

BACKGROUND

Description of the condition

There is a growing trend towards de-psychopathologisation of transgenderism (Drescher 2014; ATME 2015). There is an emerging consensus that transgenderism is not a psychiatric disorder (WPATH 2011). For instance, the 11th Revision of the International Classification of Diseases (ICD-11) (WHO 2018) no longer classifies transgenderism as a behavioural and personality disorder, but has instead drafted the term "gender incongruence" to describe gender dysphoria.

In contrast, the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM–5) (DSM-5 2013) describes gender dysphoria as a "marked incongruence between one's experienced/expressed gender and assigned gender, of at least six months duration, as manifested by at least two of the following" characteristics:

- A marked incongruence between one's experienced/expressed gender and primary and/or secondary sex characteristics (or, in young adolescents, the anticipated secondary sex characteristics);
- A strong desire to be rid of one's primary and/or secondary sex characteristics because of a marked incongruence with one's experienced/expressed gender (or, in young adolescents, a desire to prevent the development of the anticipated secondary sex characteristics);
- A strong desire for the primary and/or secondary sex characteristics of the other gender;
- A strong desire to be of the other gender (or some alternative gender different from one's assigned gender);
- A strong desire to be treated as the other gender (or some alternative gender different from one's assigned gender);
- A strong conviction that one has the typical feelings and reactions of the other gender (or some alternative gender different from one's assigned gender).

Gender dysphoria has been defined as associated with "clinically significant distress or impairment in social, occupational or other important areas of functioning" (Zucker 2016), which may lead to substantial suffering in affected people (Deutsch 2016a; Soll 2018). Gender dysphoria may result in the desire to modify one's physical and sexual characteristics to be consistent with those of the experienced gender. This process of adaptation is called transition.

The treatments applied in transition differ from those used for maintenance of the new sexual characteristics. Currently, there is uncertainty about the value of hormone therapy as a sole intervention, or when combined with surgery, for transition from male to female. This Cochrane Review specifically focuses on 'transgender women in transition from male to female,' a definition that includes biological males aiming to adapt their sexual characteristics to be consonant with those of females.

A meta-analysis that analyzed 21 studies on the prevalence of gender dysphoria (of which 12 studies contained evaluable data) estimated an overall prevalence of transgender women with gender dysphoria at 6.8 per 100,000 individuals (Arcelus 2015).

Description of the intervention

Current guidelines suggest hormonal and, if needed, surgical treatment of gender dysphoria in transgender women (WPATH 2011). Hormone therapy aims to suppress the development of, or to reverse, male attributes that have already developed. At the same time, hormones aim to develop female attributes. However, where male characteristics have already developed in adult males, such as in the bone structure of the face, hormones are not effective. Other treatments, such as surgery, would be required to change these (WPATH 2011).

The guidelines of the Endocrine Society working group suggest treatment with both oestrogens and antiandrogens (Hembree 2017). Oestrogens can be administered as either oral oestrogen, absorbed through transdermal estradiol patches, or by injection of estradiol valerate or estradiol cypionate. The application frequency differs depending on the patient's reaction to the agent and the administration regimen; it could be multiple times per day or once every two weeks. Meanwhile, antiandrogens such as spironolactone or cyproterone acetate (CPA) are commonly taken orally. Additionally, it is possible to block male puberty by treatment with gonadotropin-releasing hormone (GnRH) agonist injections (Hembree 2017).

While not every transgender woman undergoes hormone therapy in her transition, this intervention is still widely used (Hembree 2017). We know of no studies identifying the ratio of patients who undergo hormone therapy, nor do we know of studies investigating how much time passes between the start of transition (the decision to transition) and the start of hormone therapy. We are not aware of any studies on how often antiandrogens are being prescribed in addition to or instead of 17-beta-estradiol, how often they are being taken, or which kinds of androgens are in use besides CPA and spironolactone.

How the intervention might work

Several hormonal substances and combinations are used clinically for hormone therapy in transitioning women. CPA is a progestin, steroidal anti-androgen and anti-gonadotropin that blocks the receptors for testosterone (T) and dihydrotestosterone (DHT), and thereby prevents these steroidal hormones from exerting their androgenic effects. Hence, it stops processes like body hair growth, hair loss on the head, male body fat distribution and others (Figg 2010; WPATH 2011). According to the World Professional Association for Transgender Health (WPATH) guidelines, it is possible to suppress puberty with GnRH analogues or progestins such as medroxyprogesterone (WPATH 2011).

Spironolactone acts as a weak androgen receptor antagonist (Wenqing 2005). It also causes an increase in oestradiol levels (Thompson 1993), so that further virilisation is prevented and feminisation occurs (WPATH 2011).

17-beta-estradiol is used to feminise the external appearance (WPATH 2011). It binds to oestrogen receptors and thus ensures gene expression, which in turn feminises appearance (Hye-Rim 2012). In addition, estradiol suppresses gonadal testosterone production via the control systems of the hypothalamus (Hayes 2000).

Feminisation therapy aims to adapt the physical appearance and experience of the male body to that of a female body, by

inducing breast growth, softening facial features, and inducing other physical changes commonly considered to comprise a feminine appearance (WPATH 2011). For this purpose, oral or transdermal oestrogen is recommended, and therapy with oestrogen in combination with antiandrogens is most common. Co-treatment with antiandrogens minimises the required dose of oestrogen, and thereby reduces the potential risks of oestrogen identified in previous studies (Schürmeyer 1986; Prior 1989). Some antiandrogens are approved by WPATH, such as spironolactone, cyproterone acetate, GnRH analogues like goserelin, and 5-alpha-reductase inhibitors like finasteride (WPATH 2011).

Why it is important to do this review

Antiandrogens like CPA and spironolactone are prescribed to transgender women in transition by clinicians, including gynaecologists and endocrinologists (Schneider 2006; Flütsch 2015), and they are commonly considered to be valuable drugs to support transition (WPATH 2011; Hembree 2017). However, clinical evidence suggests that taking these drugs can result in adverse events; for example, CPA has significant potential for causing depression and for worsening depressive symptoms (Seal 2012). There is also some concern that CPA can lead to other psychiatric, neurological, and metabolic disorders (Griard 1978; Ramsay 1990; Oberhammer 1996; Giltay 2000; Calderón 2009; Bessone 2015). The most common adverse effects of spironolactone are hyperkalaemia, dehydration and hyponatraemia (Greenblatt 1973). Furthermore, spironolactone might have an influence on feelings of anxiety (Fox 2016).

Other studies from the 1980s and 90s reported that there were adverse effects from high-dose estradiol, but these studies used ethinyl estradiol or equine premarin (equine estradiol) instead of bioidentical 17-beta-estradiol; and used progestins, instead of bioidentical progesterone. This may have contributed to the adverse effect profile of these specific treatments (Prior 1989). Unlike the bioidentical alternatives used today (hormone preparations made from plant sources that are similar or identical to human hormones), substances administered in the past (e.g. equine oestrogens, ethinyl estradiol) were associated with more diverse adverse effects like thrombophilia, cardiovascular problems, breast and prostate cancer, as well as liver, adrenal gland and neural dysfunction (Griard 1978; Calderón 2009; Asscheman 2011). The health risks attributed to estradiol doses high enough to suppress androgens have not been found in the parenteral or transdermal application of bioidentical estradiol (Hembree 2017). Thus, it is unclear why those estradiol doses should be kept low in order to make the addition of androgen antagonists like CPA or spironolactone necessary.

In light of discussions among experts (Seal 2012; Wierckx 2014), and current recommendations for hormonal gender affirmation treatment (WPATH 2011) (which are strongly based on the values and preferences of health consumers), it is necessary to review the evidence from trials that show results for outcomes such as feminisation, satisfactory sexual function, reduced gender dysphoria, and improved quality of life (e.g. Murad 2010).

In 2017, the overall quality of evidence relating to these outcomes was classified as low (Hembree 2017). In 2011, WPATH summarised the situation as follows. "There is a need for further research on the effects of hormone therapy without surgery, and without the goal of maximum physical feminisation or masculinisation" (WPATH 2011). It is necessary to determine whether subsequent trials have provided additional evidence for efficacy, or whether there is still a lack of evidence for these desired outcomes.

OBJECTIVES

We aimed to assess the efficacy and safety of hormone therapy with antiandrogens, estradiol, or both, compared to each other or placebo, in transgender women in transition.

METHODS

Criteria for considering studies for this review

Types of studies

We aimed to include randomised controlled trials (RCTs), quasi-RCTs and controlled cohort studies.

We chose to include quasi-RCTs and cohort studies due to the low prevalence of the condition and the consequent current scarcity of RCTs (WPATH 2011).

Types of participants

We aimed to include studies that enrolled transgender women, age 16 years and over, in transition from male to female. Transitioning is defined as the process of changing one's gender profile or sexual characteristics (or both) to accord with one's sense of gender identity (WPATH 2011). Transition as a concept thus encompasses several aspects, e.g. social, psychological, or physical aspects, or a combination of these. There is consistency in the literature on when the transition begins: namely, with the decision to change a person's gender assignment (Brown 1996). However, we did not differentiate among any supposed phases of the respective types of transitions. Depending on the personal situation, the process of transition (which may include the decision to transition, gathering of information, gathering of experience, medical treatment and change of social role), can take very different periods of time, usually several months to years. Therefore, it is difficult to distinguish certain 'phases' of this process. When focusing on hormone therapy, the transition term can be more precisely defined. The transition process lasts as long as patients are in the process of changing their sexual characteristics (WPATH 2011).

We aimed to include studies with participants age 16 years and older because, according to currently applied guidelines, this is the age when patients start being treated with hormone therapy. Patients below this age are usually being treated with puberty blockers, which are outside the scope of this review (WPATH 2011).

Types of interventions

We considered studies evaluating hormone-based interventions only, excluding those that examined combined hormonal and either psychological or surgical treatments. We aimed to include studies reporting treatment with the following experimental interventions.

- Antiandrogens (cyproterone acetate or spironolactone) and estradiol
- Antiandrogens (cyproterone acetate or spironolactone) alone
- Estradiol alone

For the above interventions, we considered all types of administration: oral, sublingual, transdermal, subdermal and intramuscular. For estradiol, we also considered bioidentical 17-beta-estradiol, as well as synthetic derivatives.

We aimed to include the following comparator interventions.

- Any of the active interventions listed above
- Placebo

Although we consider placebo-controlled studies to be unethical (Bostick 2008), we made them eligible for inclusion in this review so that we could consider the evidence in its entirety. We did not consider interventions consisting purely of psychological treatment, spiritual support, or conversion therapy.

Types of outcome measures

For studies with repeated follow-up (i.e. reporting of outcomes at multiple time points), we regarded follow-up at three to six months as short term, six months to two years as medium term, and more than two years as long term (WPATH 2011).

We intended to include in the descriptive section of the review all studies that met the criteria for type of study, participants, intervention and comparator, regardless of outcomes reported or missing data.

Primary outcomes

- Quality of life (QoL) as measured by validated generic instruments, e.g. Quality of Life Inventory (QOLI) (Frisch 2005); or specific instruments, e.g. for body image, the Body Image Quality of Life Inventory (BIQLI) (Cash 2004); or for sexual life the Sexual Satisfaction Scale for Women (SSS-W) (Meston 2005).
- Satisfaction with change of male to female body characteristics, as measured with validated instruments
- Adverse events specific to hormone therapy, including serious adverse events

Secondary outcomes

- Severity of gender dysphoria/gender incongruence, e.g. as measured with the Utrecht Gender Dysphoria Scale (UGDS) (Schneider 2016)
- Measures of specific body changes, including:
  - breast size, e.g. by measurement of bust girth;
  - skin thickness, e.g. by echographic measurement (Laurent 2007);
  - skin sebum production, e.g. as measured by three-hour sebum collection with absorbent paper (Downing 1981; Giltay 2008; Ezerskaia 2016); and
  - hair growth, including hair density, diameter, growth rate and anagen/telogen ratio (Giltay 2000; Hoffmann 2013).
- Incidence or severity of depression.

We did not include surrogate outcomes, such as serum hormone levels (e.g. 17-beta-estradiol or testosterone). While these measures can help with monitoring the progress of hormone therapy, they are of little interest of themselves, especially since individuals require varying levels of these hormones to achieve a certain level of feminisation (Gooren 2017).

Search methods for identification of studies

Electronic searches

We searched the following electronic databases for relevant trials up to 19 December 2019 with no restrictions based on language of publication, date of publication, or publication status:

- MEDLINE via PubMed
- Cochrane Central Register of Controlled Trials (CENTRAL)
- Embase
- Biosis Preview
- PsycINFO
- PSYNDEX

Our search strategy is outlined in Appendix 1. We have successfully tested the screening methods for abstracts and titles.

Searching other resources

Had we identified any eligible studies through the electronic searches above we would have searched the reference lists of these in order to find additional relevant studies. We also searched the scientific abstracts of the last two meetings of each of the following organisations:

- American Association of Clinical Endocrinologists
- American Society of Andrology
- Berufsverband der deutschen Endokrinologen (Professional Association of the German Endocrinologists)
- Berufsverband der Frauenärzte e.V. (Professional Association of the Gynaecologists)
- Dachverband Reproduktionsbiologie und Medizin e.V. (Federal Association Reproductive Biology and Medicine)
- Deutsche Gesellschaft für Endokrinologie (German Society for Endocrinology)
- Deutsche Gesellschaft für Gynäkologie und Geburtshilfe (German Society for Gynaecology and Obstretics)
- Endocrine Society
- European Society of Gynaecological Oncology
- European Thyroid Association
- Nordrhein-Westfälische Gesellschaft für Endokrinologie und Diabetologie (North Rhine-Westphalian Society for Endocrinology and Diabetology)
- Royal College of Obstetricians and Gynaecologists
- Society for Endocrinology
- Society for Gynaecologic Investigation

We also searched the following grey literature databases:

- The New York Academy of Medicine Grey Literature Report (www.greylit.org/)
- OAIster (www.oclc.org/oaister.en.html)
- OpenGrey (www.opengrey.eu/)

Finally, in order to identify completed but unpublished or ongoing studies, we searched the following trial registries.

- ClinicalTrials.gov (www.clinicaltrials.gov/)
- metaRegister of Controlled Trials (mRCT; www.controlledtrials.com/mrct/)

- World Health Organization (WHO) International Clinical Trials Registry Platform (ICTRP) Search Portal (www.who.int/trialsearch/)
- Drugs@FDA (www.accessdata.fda.gov/scripts/cder/drugsatfda/)
- European Public Assessment Reports (EPAR; www.ema.europa.eu/ema/index.jsp?curl=pages/medicines/landing/epar_search.jsp)

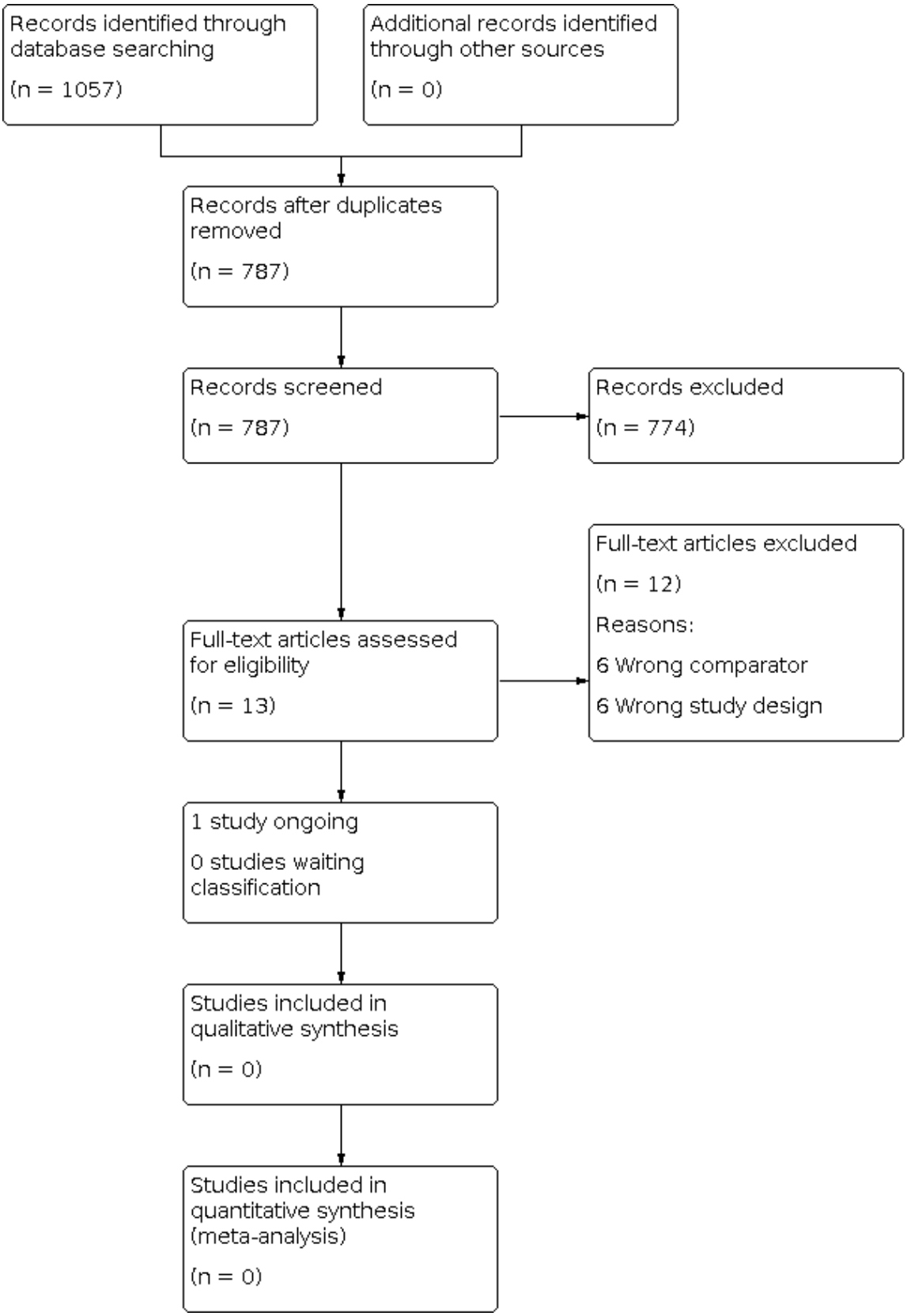
We contacted fifteen manufacturers of hormonal agents and experts in the field to identify unpublished or ongoing trials.

Data collection and analysis

Selection of studies

We used the reference management tool Covidence to identify and remove potential duplicate records of relevant studies (www.covidence.org). Two review authors (AKU and MHE) independently scanned titles and abstracts of the remaining records to compile a list of potential papers to potentially be included in the review. After this, the same review authors investigated the references in detail (as full text articles or matched records to studies), and categorised these as 'included studies,' 'excluded studies,' 'studies awaiting classification' and 'ongoing studies.' We executed this task in accordance with the criteria provided in the *Cochrane Handbook for Systematic Reviews of Interventions* (Higgins 2011a). If there had been discrepancies or if a consensus could not be reached, a third review author would have adjudicated (CHA). There were no disagreements that could not be thus resolved. Had this been the case, we would have designated the study as 'awaiting classification' and contacted the study authors for clarification. We listed studies excluded during the full text review stage, and documented the reasons for exclusion in Characteristics of excluded studies. We included an adapted PRISMA flow diagram outlining the study selection process (Moher 2009) (Figure 1).

Figure 1. Study Flow Diagram



Data extraction and management

If we had found relevant studies, two review authors (AKU and MHE) would have extracted data from all studies deemed eligible for inclusion independently, with the help of a standardized data extraction form that would have been pilot tested according to Chapter 7 of the *Cochrane Handbook* (Higgins 2011a). We have used Google Spreadsheets to manage all data gathered.

We would have collected data on the following items:

- General information on the study: first author, date of publication, study dates, publication type (full text article, abstract, unpublished), citation.
- Study methods: study design (e.g. parallel, factorial), number of study arms, study setting (single institution, multi-centre national, multi-centre international), study location, and length of follow-up.
- Participant characteristics: study inclusion/exclusion criteria, age (mean/median with range), ethnic distribution, number of participants randomised and included in analysis, participants lost to follow-up.
- Interventions: type of hormonal agents (for example CPA, estradiol, progesterone, spironolactone), dose, administration route, dosing schedule and any other associated therapies. We would have extracted data on the sample size for each intervention group.
- Outcomes: definition and method of assessment for each outcome (including the adverse event classification system used in individual studies), as well as any relevant subgroups. We would have extracted the number of events and participants per treatment group for dichotomous outcomes. We would also extract the mean, standard deviation or median and range, and number of participants per treatment group for continuous outcomes.
- Study funding sources.
- Declarations of potential conflicts of interest reported by study authors.

For each included study, we would have extracted the outcome data relevant for this review, and which would be required for the calculation of summary statistics and measures of variance. If there had been disagreements, we would have resolved them by discussion. If necessary, we would have consulted a third review author (CHA). We provided key information about potentially relevant ongoing studies, including trial identifiers, in the table of Characteristics of ongoing studies. We would have attempted to contact authors of included studies to obtain missing key data if needed.

Assessment of risk of bias in included studies

If relevant studies had been found, two review authors (AKU and MHE) would have examined all included studies to assess risk of bias (assessment of methodological quality) independently. We would have used the Cochrane 'Risk of bias' tool for assessing risk of bias in RCTs, as described in the *Cochrane Handbook* (Higgins 2011b). We would have resolved disagreements by consensus or by consulting a third review author (CHA). Our summary judgement would have included a rating (low, high or unclear risk of bias) for each domain (Higgins 2011b). We would have assessed the risk of bias for the following domains:

- Random sequence generation
- Allocation concealment
- Blinding of participants and personnel
- Blinding of outcome assessment
- Incomplete outcome data
- Selective reporting
- Other bias

We would have evaluated the risks of performance bias (blinding of participants and personnel) and detection bias (blinding of outcome assessment) separately for each outcome.

For any relevant cohort studies we would have used the ROBINS-I tool to assess risk of bias (Sterne 2016). We would have assessed each individual study in accordance with the guidance, documenting the results using a spreadsheet and providing details in 'Risk of bias' tables. We would have documented the reasons for our judgements, and would have included relevant quotations from the full-text articles or from information about the study provided by authors in the notes section of the 'Risk of bias' tables. We would have summarised the risk of bias across domains for each primary outcome in every included study, as well as across studies and domains for each primary outcome.

Measures of treatment effect

Dichotomous data

We planned to summarise dichotomous data using risk ratios (RRs), reported with 95% confidence intervals (CIs).

Continuous data

For continuous outcomes with a standard measure, we would have summarised the obtained data as mean differences (MDs) with 95% CIs. For continuous outcomes without a standard measure, we would have summarised data as standardized mean differences (SMDs) with 95% CIs. Alternatively, if the mean value and variance were missing, we would have estimated them using the methods described in Hozo 2005, which allows estimations for mean value and variance of a sample when only the median, range and size of the sample are known. We would also have considered the guidance in the *Cochrane Handbook* where appropriate (Higgins 2011c).

Unit of analysis issues

We planned to treat recurring events in individual participants as single events occurring in one participant (e.g. three episodes of major depressive disorder in one participant would have been recorded as one participant with major depressive disorder). We did not expect to include studies with interventions delivered at the cluster level.

Dealing with missing data

For studies with missing data, we would have followed the recommendations of the *Cochrane Handbook* (Higgins 2011d). We would have collected dropout rates for each study group and would have reported these in the 'Risk of bias' table. Our preferred option would have been to contact study authors in cases of missing data or statistics that were not due to participant dropout (e.g. missing statistics such as standard deviation (SD)). If missing outcome data were not provided, then we would have attempted to impute

data where possible and appropriate, and conduct sensitivity analyses to assess the effect of this on the analysis. However, where imputation is not appropriate, we would not have included the study in the respective meta-analysis, and would have discussed the potential impact of this in the text of the review. In the case of participants lost to follow-up, we would have performed meta-analyses on an intention-to-treat basis. We would have performed sensitivity analyses, excluding studies with missing outcome data, to evaluate the impact of missing data. We would have discussed the potential impact of missing data on review findings in the 'Discussion' section of the full review, using a summary table if appropriate.

Assessment of heterogeneity

We would have compared the characteristics of included studies to identify heterogeneity of content or methodology, and to determine the feasibility of performing a meta-analysis. We would have deemed meta-analyses unsuitable in cases where there was substantial content-related or methodological heterogeneity across studies. Instead, we would have used a narrative approach to data synthesis. Had meta-analyses been deemed appropriate, we would have assessed statistical heterogeneity by visually inspecting the scatter of individual study effect estimates on forest plots and by calculating the I^2 statistic (Higgins 2011c), which gives the percentage of variability in effect estimations that can be attributed to heterogeneity rather than to chance. We would have considered an I^2 of more than 50% to represent substantial heterogeneity. In the case of statistical heterogeneity, we would have conducted the prespecified subgroup and sensitivity analyses described below to investigate the source.

Assessment of reporting biases

If we had included 10 or more studies that investigated the same outcome, we would have used funnel plots to assess small-study effects and publication bias. Given that several explanations are possible for funnel plot asymmetry, we would have interpreted results carefully (Sterne 2011).

Data synthesis

Had we identified any eligible studies, we would have provided a narrative summary of the included studies. We would also have conducted meta-analyses of RCTs for all relevant outcomes, where possible, using data from studies that 1) compared the actual hormone therapy-relevant agents or combinations of agents to placebo, and 2) compared the actual hormone therapy-relevant agents or combinations of agents to other hormone therapy-agents or combinations of agents. Studies comparing two variations on the intervention would have been pooled separately to studies comparing the intervention to placebo. However, if there had been significant variability in the definition of outcomes across trials, we would have decided not to pool data.

Had we conducted meta-analyses, we would have used the Mantel-Haenszel approach to combine dichotomous data and calculate RRs with 95% CIs (Higgins 2011c). For continuous outcomes (e.g. quality of life) we would have calculated MDs or SMDs, with 95% CIs, using the inverse variance approach. Had studies reported the same outcome measure but some studies had reported data on the change from baseline (e.g. mean values and standard deviations) and others for final measurements of outcomes, they would have been placed in subgroups in the meta-analysis and pooled according to guidance in the *Cochrane Handbook* (Higgins 2011c).

For meta-analyses, we would have used a random-effects model, expecting the true effects to be related, but not the same, across all studies. We would have interpreted random-effects meta-analyses with due consideration of the whole distribution of effects, ideally by presenting a prediction interval (Higgins 2009). A prediction interval specifies a predicted range for the true treatment effect in an individual study (Riley 2011). In addition, we would have performed statistical analyses according to the statistical guidelines contained in the *Cochrane Handbook* (Higgins 2011c).

We would have summarised outcome data from cohort studies (e.g. change scores) narratively.

Subgroup analysis and investigation of heterogeneity

Wherever possible, we would have considered subgroup analyses that are structured by the following characteristics.

- Type of application of intervention (oral, transdermal, intramuscular, subcutaneous)
- Orchiectomy before or during hormone therapy

The justification for these analyses is as follows. Pharmacokinetic mechanisms lead to significant differences in the absorption and metabolism of an active substance depending on the type of application. Therefore, we would, if possible, have formed appropriate subgroups based on the application method of the intervention. Also, patients who have undergone an orchiectomy could have different outcomes than those patients without orchiectomy (Defreyne 2017).

Sensitivity analysis

We would have conducted sensitivity analyses to investigate any potential effect of removing studies judged to be at high risk of bias from meta-analyses. We would have classified studies as being at high risk of bias overall if one or more domains were judged to be at high risk. If appropriate, we would also have conducted sensitivity analyses excluding studies with missing outcome data, or where missing data have been imputed by the review author team. We would also have conducted a sensitivity analysis to compare a fixed-effect model to a random effects model where the studies in a meta-analysis appear more homogeneous than expected.

Summary of findings and assessment of the certainty of the evidence\

Following standard Cochrane methodology, had we identified any included studies, we would have created a 'Summary of findings' table for all three primary outcomes. Also following standard Cochrane methodology, we would have used the five GRADE considerations (risk of bias, consistency of effect, imprecision, indirectness and publication bias) to assess the quality of the body of evidence for each outcome, and to draw conclusions about the quality of evidence within the text of the review.

RESULTS

Description of studies

Results of the search

We conducted our searches on 18 January 2019 and updated them on 19 December 2019. Through the database searches, we identified a total of 1057 references. After removing duplicates, we screened the titles and abstracts of 787 references. Through this screening, we identified 13 studies to assess as full text articles. We fully inspected these articles, and excluded 12 studies. The remaining study was still ongoing. Therefore, we did not include any studies in this review (Figure 1).

Of the manufacturers and experts in the field whom we contacted,15 responded but did not report any additional studies.

Included studies

None of the reports retrieved met the inclusion criteria for this review. Suggestions for future studies are given in Table 1

Excluded studies

We excluded all 12 of the full-text articles that we had assessed for eligibility, either because they used an ineligible comparator or because they used an ineligible study design. See Characteristics of excluded studies for further details.

Ongoing studies

We identified one ongoing RCT in Thailand, comparing spironolactone with CPA (Krasean 2019). This study started in April 2019. We describe this study in Characteristics of ongoing studies.

Risk of bias in included studies

As no studies met the inclusion criteria, it was not possible to assess risk of bias.

Effects of interventions

As no studies met the inclusion criteria, we were unable to calculate any effects of the interventions.

DISCUSSION

Summary of main results

No study met the inclusion criteria for this review. A total of 13 potentially eligible studies were identified, but ultimately all but one was excluded after we assessed the full text articles. The one remaining RCT is ongoing, and we are awaiting its publication (Krasean 2019). We conducted a comprehensive search to identify eligible studies for inclusion in this review. Despite more than four decades of ongoing efforts to improve the quality of hormone therapy for women in transition, we found that no RCTs or suitable cohort studies have yet been conducted to investigate the efficacy and safety of hormonal treatment approaches for transgender women in transition.

Overall completeness and applicability of evidence

The evidence is incomplete because no studies met the inclusion criteria for the review. This lack of studies shows a gap between current clinical practice and clinical research, which has been repeatedly emphasised (Hembree 2009; Hembree 2017). If hormone therapy is highly valued in the treatment of gender dysphoria (Hembree 2009; WPATH 2011; Hembree 2017), then this raises the question: why are there no RCTs or appropriate cohort studies for this clinical condition? There is also an ethical need for research into the efficacy and safety of hormone therapy, particularly comparing combination therapy with CPA/estradiol and spironolactone/estradiol to monotherapy with estradiol alone. In view of the reported but rather alarming side-effect profiles of CPA and spironolactone in other populations (De Bastos 2014; Khan 2016; PG12 2019), long-term clinical studies that aim to achieve adequate outcomes are urgently needed for the population of transgender women in transition.The lack of reliable data on hormone therapy for transitioning transgender women should encourage the development of well-planned RCTs and cohort studies to evaluate widespread empirical practice in the treatment of gender dysphoria.

The most common reason for the exclusion of studies from this review was the lack of a control group. We excluded some studies because they did not meet the eligibility requirements for study design (e.g. case series or case-control studies). Further, interventions were not clearly defined.

Among guideline developers in the field of transgender medicine, it has been discussed in recent years why the available evidence remains limited (Deutsch 2016a Reilly 2019). Deutsch 2016a has identified three main reasons, which they believe have hindered the development of evidence based healthcare guidelines. Firstly, a lack of research funding and institutional stigma means that the evidence currently centres around less robust study designs, such as retrospective studies, case series, and individual case reports (Bockting 2016 Reisner 2016a); secondly variation in the collection of gender identity data in observational data sets makes it difficult to identify relevant populations and monitor their health outcomes (Deutsch 2013 Bauer 2009); and finally, academic programmes focused on transgender medicine are in their infancy and few exist (Reisner 2016b), meaning there is a general lack of research and training on this topic.

Against this background, methodological problems such as inconsistent and missing comparison groups, uncontrolled confounding factors, small sample size, short follow-up time and difficulties in recording and evaluating a broad spectrum of health outcomes (physical and mental health, social functioning and QoL) have become apparent in hormone therapy (Deutsch 2016b). The performance of RCTs is controversial, especially with regard to placebo studies, and ethical and methodological objections have been raised (e.g. violation of the principle of equipoise, Miller 2003). However, the positive research potential of active-controlled RCTs is acknowledged, in order to compare different types, dosages and methods of administration of active treatments. Overall, there is a trend in the discussion to favour not only RCTs and quasi-RCTs, but also high-quality cohort studies conducted in a network of health centres, hospitals and practices (Deutsch 2016a; Deutsch 2016b).

Quality of the evidence

We could not appraise the quality of the evidence because no studies met our review's inclusion criteria.

Potential biases in the review process

We consider our search to have been consistent and comprehensive (including the fifteen contacts with manufacturers and experts in the field). At each stage, the review authors independently applied the inclusion criteria before comparing their judgements. Reliability testing was performed in the screening phase. Even though we were unable to test for publication bias, we think it is unlikely that there are studies that have been conducted but remained unpublished. The experts in the field we interviewed believed that there was a general lack of research activity by treatment manufacturers, and considered it very likely that no phase IV studies have ever been conducted in this population. For example, one expert stated that there was probably "nothing to be kept secret."

Agreements and disagreements with other studies or reviews

There are currently no systematic reviews in the Cochrane Library that evaluate the effectiveness of hormone therapy for transgender women in transition, nor are there systematic reviews that evaluate the clinical and economic impact of hormone therapy on transgender women in transition. The Endocrine Society's 2009 and 2017 guidelines addressed endocrine treatment of gender-dysphoric/gender-incongruent persons (Hembree 2009; Hembree 2017). The literature search included in these guidelines did not identify any RCTs of hormone therapy in transitioning transgender women. In the context of the preparation of UK National Health Service (NHS) guidelines (PG12 2019), the NHS Guideline Panel also found no RCTs. However, PG12 2019 includes a recommendation for the prescription of hormone therapy for transitioning transgender women.

Of the potentially relevant studies we excluded, some reported on relevant questions. Asscheman 2011 focused on the important outcome of mortality. Fisher 2016 investigated the important relationship between hormone therapy-related body changes and psychobiological well-being. Giltay 2000 focused on body related outcomes such as hormone therapy's effects on the skin (hair growth rate, density, and shaft diameter by image analysis; and sebum production). Toorians 2003 focused on the outcomes of different interventions (estradiol alone compared with combination therapy estradiol and antiandrogens). Miles 2006 was based on a cross-over design with the intention of comparing groups of individuals on and off oestrogen. Due to the reported deficits, we excluded these studies, although they addressed important questions.

AUTHORS' CONCLUSIONS

Implications for practice

We found insufficient evidence to determine the efficacy or safety of hormonal treatment approaches (estradiol alone or in combination with cyproterone acetate or spironolactone) for transgender women in transition. The evidence is very incomplete, demonstrating a gap between current clinical practice and clinical research.

Implications for research

This systematic review has shown that well-designed, sufficiently robust randomised controlled trials (RCTs) and controlled-cohort studies do not exist, and are needed, to assess the benefits and harms of hormone therapies (used alone or in combination) for transgender women in transition. The following questions should be addressed via RCTs and cohort studies:

1. What are the short-, medium-, and long-term effects (including adverse effects, benefits, and prognoses) of estradiol therapy alone, as opposed to combination therapy using estradiol together with cyproterone acetate or spironolactone?
2. What is the short-, medium-, and long-term clinical efficacy of hormone therapy when applied orally, transdermally, and intramuscularly?

Table 1 presents design components that we suggest could be used in future studies. Studies should be structured and reported according to the CONSORT Statement or the STROBE Statement in order to improve the quality of reporting on efficacy and to obtain better reports on harms in clinical research (von Elm 2007; Schulz 2010). There is an urgent need for research in this area, not least for ethical reasons.

We will include non-controlled cohort studies in the next iteration of this review, as this review has demonstrated that this is the highest quality evidence currently available in the field. We will take methodological limitations into account when doing so.

ACKNOWLEDGEMENTS

We would like to acknowledge the support of Dr. Jonathan Livingstone-Banks, Dr. Nicola Lindson, and Dr. Paul Aveyard from the Tobacco Addiction Group, as well as Dr. Erik von Elm from Cochrane Switzerland, whom we consulted in preparing this review.

We greatly appreciate Dr. Alissa Jones Nelson's support in reviewing the spelling and grammar of this review. We also gratefully acknowledge peer review comments from Igor Grabovac, Department of Social and Preventive Medicine, Centre for Public Health, Medical University of Vienna, Vienna, Austria and Dr. Barbara Nussbaumer-Streit, Cochrane Austria, Department for Evidence-based Medicine and Evaluation, Danube-University Krems, Krems, Austria, and consumer review comments from Sarah Stephenson-Hunter.

REFERENCES

References to studies excluded from this review

Asscheman 2011 *{published data only}*

Asscheman H, Giltay EJ, Megens JA, de Ronde WP, van Trotsenburg MA, Gooren LJ. A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones. *European Journal of Endocrinology* 2011;**164**(4):635-42. [DOI: 1530/EJE-10-1038]

Colizzi 2015 *{published data only}***10.1016/j.jpsychores.2015.02.001**

Colizzi M, Costa R, Scaramuzzi F. Concomitant psychiatric problems and hormonal treatment induced metabolic syndrome in gender dysphoria individuals: A 2 year follow-up study. *Journal of Psychosomatic Research* 2015;**78**:399–406. [DOI: 10.1016/j.jpsychores.2015.02.001]

Fighera 2018 *{published data only}***10.1111/cen.13607**

Fighera TM, da Silva E, Lindenau JD, Spritzer PM. Impact of cross-sex hormone therapy on bone mineral density and body composition in transwomen. *Clinical Endocrinology* 2018;**88**(6):856-862. [DOI: 10.1111/cen.13607] [PMID: 29630732]

Fisher 2014 *{published data only}***10.1111/jsm.12413**

Fisher AD, Castellini G, Bandini E, Casale H, Fanni E, Benni L, et al. Cross-sex hormonal treatment and body uneasiness in individuals with gender dysphoria. *International Society for Sexual Medicine* 2014;**11**:709–19.

Fisher 2016 *{published data only}***10.1210/jc.2016-1276**

Fisher AD, Castellini G, Ristori J, Casale H, Cassioli E, Sensi C, et al. Cross-sex hormone treatment and psychobiological changes in transsexual persons: two-year follow-up data. *The Journal of Clinical Endocrinology and Metabolism* 2016;**101**:0000-0000.

Giltay 2000 *{published data only}*

Giltay EJ, Gooren L. Effects of sex steroid deprivation/administration on hair growth and skin sebum production in transsexual males and females. *The Journal of Clinical Endocrinology & Metabolism* 2000;**85**(8):2913-21. [DOI: 10.1210/jc.85.8.2913]

Haraldsen 2005 *{published data only}*

Haraldsen IR, Egeland T, Haug E, Finset A, Opjordsmoen S. Cross-sex hormone treatment does not change sex-sensitive cognitive performance in gender identity disorder patients. *Psychiatry Research* 2005;**137**:161-74. [10.1016/j.psychres.2005.05.014]

Haraldsen 2007 *{published data only}***10.1016/j.yhbeh.2007.05.012**

Haraldsen IR, Haug E, Falch J, Egeland T, Opjordsmoen S. Cross-sex pattern of bone mineral density in early onset gender identity disorder. *Hormones and Behavior* 2007;**52**:334-43. [DOI: 10.1016/j.yhbeh.2007.05.012]

Miles 2006 *{published data only}***10.1016/j.yhbeh.2006.06.008**

Miles C, Green R, Hines M. Estrogen treatment effects on cognition, memory and mood in male-to-female transsexuals. *Hormones and Behavior* 2006;**50**:708-17. [DOI: 10.1016/j.yhbeh.2006.06.008]

Schlatterer 1998 *{published data only}*

Schlatterer K, Auer DP, Yassouridis A, Von Werder K, Stalla GK. Transsexualism and osteoporosis. *Archives of Sexual Behavior* 1998;**27**(5):475-92. [0004-0002/98/1000-0475]

Toorians 2003 *{published data only}*

Toorians AW, Thomassen MCLGD, Zweegman S, Magdeleyns EJP, Tans G, Gooren L, et al. Venous thrombosis and changes of hemostatic variables during cross-sex hormone treatment in transsexual people. *The Journal of Clinical Endocrinology & Metabolism* 2003;**88**(12):5723-29. [DOI: 10.1210/jc.2003-030520]

Van Goozen 1995 *{published data only}*

Van Goozen SHM. Gender differences in behaviour: activating effects of cross-sex hormones. *Psychoneuroendocrinology* 1995;**20**(4):343-63. [DOI: 10.1016/0306-4530%2894%2900076-X]

References to ongoing studies

Krasean 2019 *{published data only}*

TCTR20190404001. Anti-androgenic effects comparison between Cyproterone acetate and Spironolactone in transgender women: a randomized controlled trial. Thai Clinical Trials Registry 2019. [THAI CLINICAL TRIALS REGISTRY: TCTR20190404001]

Additional references

Arcelus 2015

Arcelus J, Bouman WP, Van Den Noortgate W, Claes L, Witcomb G, Fernandez-Aranda F. Systematic review and meta-analysis of prevalence studies in transsexualism. *European Psychiatry* 2015;**30**(6):807-15.

ATME 2015

Aktion Transsexualität und Menschenrecht eV (ATME). Alternative recommendations for treatment in the presence of so-called "sex/gender variance". Medicine and psychotherapy without gender stereotyping. [STUTTGARTER ERKLÄRUNG - Alternative Handlungsempfehlungen bei geschlechtlichen Normvariationen]. In: In: v. Schreiber G editor(s). Transsexualität in Theologie und Neurowissenschaften - Ergebnisse, Kontroversen, Perspektiven. Vol. **1**. Berlin: De Gruyter, 2015:77-8. [ISBN: 978-3110440805]

Bauer 2009

Bauer GR, Hammond R, TraversR, Kaay M, Hohenadel KM, Boyce M. “I don't think this is theoretical; this is our lives”: how erasure impacts health care for transgender people. *Journal of the Association of Nurses in AIDS Care* 2009;**20**(5):348-61. [DOI: 10.1016/j.jana.2009.07.004]

Bessone 2015

Bessone F, Lucena MI, Roma MG, Stephens C, Medina-Cáliz I, Frider B, et al. Cyproterone acetate induces a wide spectrum of acute liver damage including corticosteroid-responsive hepatitis: report of 22 cases. *Liver International: Official Journal of the International Association for the Study of the Liver.* 2015;**36**(2):302-10. [DOI: 1111/liv.12899]

Bockting 2016

Bockting W, Coleman E, Deutsch MB, Guillamon A, Meyer I, Meyer III W, et al. Adult development and quality of life of transgender and gender nonconforming people. *Current Opinion in Endocrinology, Diabetes, and Obesity* 2016;**23**(2):188. [DOI: 10.1097/MED.0000000000000232]

Bostick 2008

Bostick NA, Sade R, Levine MA, Stewart Jr DM. Placebo use in clinical practice: report of the American Medical Association Council on Ethical and Judicial Affairs. *The Journal of Clinical Ethics* 2008;**19**(1):58-61. [PMID: 18552054]

Brown 1996

Brown ML, Rounsley CA. True Selves: Understanding Transsexualism - For Families, Friends, Coworkers, and Helping Professionals. 1 edition. Vol. **1**. San Francisco: Jossey-Bass, 1996. [ISBN: 0-7879-6702-5]

Calderón 2009

Calderón GD, Bratoeff E, Ramiréz LE, Osnaya BN, Garcia AR, Barragán MG, et al. Effects of two new steroids and cyproterone on some biomarkers of oxidative stress and serotonergic system on rat prostate and brain. Andrologie Feb 2009;**41**(1):29-34. [DOI: 1111/j.1439-0272.2008.00886.x]

Cash 2004

Cash TF, Jakatdar TA, Williams EF. The body image quality of life inventory: further validation with college men and women. *Body Image* 2004;**1**(3):279–87. [DOI: 1016/S1740-1445(03)00023-8]

De Bastos 2014

de Bastos M, Stegeman B, Rosendaal F. Combined oral contraceptives: venous thrombosis. *Cochrane Database of Systematic Reviews* 2014;**3**(Issue ID 2351):Art. No.: CD010813. [DOI: 10.1002/14651858.CD010813.pub2.] [PMID: 24590565]

Defreyne 2017

Defreyne J, Nota N, Pereira C, Schreiner T, Fisher AD, den Heijer M, et al. Transient elevated serum prolactin in trans women is caused by cyproterone acetate treatment. *LGBT Health* 2017;**4**(5):328-36. [DOI: 1089/lgbt.2016.0190]

Deutsch 2013

Deutsch MB, Green J, Keatley J, Mayer G, Hastings J, Hall AM, et al. Electronic medical records and the transgender patient: recommendations from the World Professional Association for Transgender Health. *Journal of the American Medical Informatics Association* 2013;**20**(4):700-3. [10.1136/amiajnl-2012-001472]

Deutsch 2016a

Deutsch M. Guidelines for the primary and gender-affirming care of transgender and gender nonbinary people. Center of Excellence for Transgender Health June 17th 2016;**2nd Edition**.

Deutsch 2016b

Deutsch MB, Radix A, Reisner S. What's in a guideline? Developing collaborative and sound research designs that substantiate best practice recommendations for transgender health care. *AMA Journal of Ethics* 2016;**18**(11):1098-106. [DOI: 10.1001/journalofethics.2016.18.11.stas1-1611]

Downing 1981

Downing DT, Stewart ME, Strauss JS. Estimation of sebum production rates in man by measurement of the squalene content of skin biopsies. *Journal of Investigative Dermatology* 1981;**77**(4):358-60. [PMID: 7276619]

Drescher 2014

Drescher J. Controversies in gender diagnoses. *LGBT Health* 2014;**1**(1):10-14. [DOI: 1089/lgbt.2013.1500]

DSM-5 2013

American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5), Fifth edition. 1 edition. Vol. **1**. Göttingen: Hogrefe, 2013. [ISBN: 9783801725990]

Ezerskaia 2016

Ezerskaia A, Pereira SF, Urbach HP, Verhagen R, Varghese B. Infrared spectroscopic measurement of skin hydration and sebum levels and comparison to corneometer and sebumeter. *Proceedings SPIE* 2016;**9887**(98872G):552-6. [DOI: 1117/12.2225434]

Figg 2010

Figg W, Chau CH, Cindy H, Small EJ. Drug Management of Prostate Cancer. 1 edition. New York: Springer, 2010. [ISBN: 978-1-60327-829-4]

Flütsch 2015

Flütsch N. Endocrinological treatment of the gender dysphoria in people with gender incongruence. [Endokrinologische Behandlung der Geschlechtsdysphorie bei Menschen mit Geschlechtsinkongruenz]. *Journal of Clinical Endocrinology and Metabolism* 2015;**8**(2):42-8.

Fox 2016

Fox LC, Davies DR, Scholl JL, Watt MJ, Forster GL. Differential effects of glucocorticoid and mineralocorticoid antagonism on anxiety behavior in mild traumatic brain injury. *Behavioural Brain Research* 2016;**312**:362-5. [DOI: 1016/j.bbr.2016.06.048]

Frisch 2005

Frisch MB, Clark MP, Rouse SV, Rudd MD, Paweleck JK, Greenstone A, et al. Predictive and treatment validity of life satisfaction and the quality of life inventory. *Assessment* 2005;**12**(1):66-78. [DOI: 1177/1073191104268006]

Giltay 2008

Giltay EJ, Bunck MC, Gooren L, Zitman FG, Diamant M, Teerlink T. Effects of sex steroids on the neurotransmitter-

specific aromatic amino acids phenylalanine, tyrosine, and tryptophan in transsexual subjects. *Neuroendocrinology* 2008;**88**(2):103-10. [DOI: 1159/000135710]

Gooren 2017

den Heijer M, Bakker A, Gooren L. Long term hormonal treatment for transgender people. *The BMJ* 2017;**359**(j5027):n/a. [DOI: 1136/bmj.j5027]

Greenblatt 1973

Greenblatt DJ, Koch-Weser J. Adverse reactions to spironolactone. *JAMA* 1973;**225**(1):40-3. [DOI: 1001/jama.1973.03220280028007]

Griard 1978

Griard J, Bühler U, Zuppinger K, Haas HG, Staub JJ, Wyss HI. Cyproterone acetate and ACTH adrenal function. *The Journal of Clinical Endocrinology and Metabolism* 1978;**47**(3):581-6. [DOI: 1210/jcem-47-3-581]

Hayes 2000

Hayes FJ, Seminara SB, Decruz S, Boepple PA, Crowley WF Jr. Aromatase inhibition in the human male reveals a hypothalamic site of estrogen feedback. *The Journal of Clinical Endocrinology & Metabolism* 2000;**85**(9):3027-35. [DOI: 0021-972X/00/$03.00/0]

Hembree 2009

Hembree WC, Cohen-Kettenis P, Delemarre-van de Waal HA, Gooren LJ, Meyer III WJ, Spack NP, et al. Endocrine treatment of transsexual persons: an endocrine society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism* 2009;**94**(9):3132–54. [DOI: 10.1210/jc.2009-0345]

Hembree 2017

Hembree WC, Cohen-Kettenis P, Delemarre-van de Waal HA, Gooren L, Hannema SE, Meyer III WJ, et al. Endocrine treatment of transsexual persons: an Endocrine Society clinical practice guideline. *The Journal of Clinical Endocrinology & Metabolism* 2017;**102**(11):3869-903. [DOI: 1210/jc.2017-01658]

Higgins 2009

Higgins JPT, Thompson SG, Spiegelhalter DJ. A re-evaluation of random-effects meta-analysis. Journal of the Royal Statistical Society: Series A (Statistics in Society) 2009;**172**(1):137-59. [DOI: 172(1):137–159]

Higgins 2011a

Higgins JPT, Deeks JJ. Chapter 7: Selecting studies and collecting data. In: Higgins JPT, Deeks JJ, editor(s), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Higgins 2011b

Higgins JPT, Deeks JJ. Chapter 8: Assessing risk of bias in included studies. In: Higgins JPT, Altman DG, Sterne JAC, editor(s), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Higgins 2011c

Higgins JPT, Deeks JJ. Chapter 9: Analysing data and undertaking meta-analyses. In: Deeks JJ, Higgins JPT, Altman DG, editor(s), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Higgins 2011d

Higgins JPT, Deeks JJ. Chapter 16: General principles for dealing with missing data. In: Higgins JPT, Deeks JJ, Altman DG, editor(s), Cochrane Handbook for Systematic Reviews of Interventions Version 5.1.0 (updated March 2011). The Cochrane Collaboration, 2011. Available from handbook.cochrane.org.

Hoffmann 2013

Hoffman R. TrichoScan: a novel tool for the analysis of hair growth in vivo. *Journal of Investigative Dermatology Symposium Proceedings* 2003;**8**(1):109-15. [DOI: 1046/j.1523-1747.2003.12183.x]

Hozo 2005

Hozo SP, Djulbegovic B, Iztok I. Estimating the mean and variance from the median, range, and the size of a sample. BMC Medical Research Methodology 2005;**13**(5). [DOI: 1186/1471-2288-5-13]

Hye-Rim 2012

Hye-Rim L, Tae-Hee K, Kyung-Chul C. Functions and physiological roles of two types of estrogen receptors, ERα and ERβ, identified by estrogen receptor knockout mouse. *Laboratory Animal Research* 2012;**28**(2):71-6. [DOI: 5625/lar.2012.28.2.71]

Khan 2016

Khan O, Mashru A. The efficacy, safety and ethics of the use of testosterone-suppressing agents in the management of sex offending. *Current Opinion in Endocrinology, Diabetes and Obesity* 2016;**23**(3):271-8. [DOI: 10.1097/MED.0000000000000257]

Laurent 2007

Laurent A, Mistretta F, Bottigioli D, Dahel K, Goujon C, Nicolas JF, et al. Echographic measurement of skin thickness in adults by high frequency ultrasound to assess the appropriate microneedle length for intradermal delivery of vaccines. Vaccine 2007;**25**(34):6423-30. [DOI: 25(34):6423-6430]

Meston 2005

Meston C, Trapnell P. Development and validation of a five-factor sexual satisfaction and distress scale for women: the Sexual Satisfaction Scale for Women (SSS-W). *The Journal of Sexual Medicine* 2005;**2**(1):66-81.

Miller 2003

Miller FG, Brody H. A critique of clinical equipoise: therapeutic misconception in the ethics of clinical trials. *Hastings Center Report* 2003;**33**(3):19-28. [PMID: 12854452]

Moher 2009

Moher D, Liberati A, Tetzlaff J, Altman DG. The PRISMA Group (2009). Preferred reporting items for systematic

reviews and meta analyses: The PRISMA statement. Annals of Internal Medicine 2009;**151**(4):264-9. [DOI: 7326/0003-4819-151-4-200908180-00135]

Murad 2010

Murad MH, Elamin MB, Garcia MZ, Mullan RJ, Murad A, Erwin PJ, et al. Hormonal therapy and sex reassignment:a systematic review and meta-analysis of quality of life and psychosocial outcomes. Clinical Endocrinology 2010;**72**(2):214-31. [DOI: 1111/j.1365-2265.2009.03625.x]

Oberhammer 1996

Oberhammer F, Nagy P, Tiefenbacher R, Fröschl G, Bouzahzah B, Thorgeirsson SS, et al. The antiandrogen cyproterone acetate induces synthesis of transforming growth factor beta 1 in the parenchymal cells of the liver accompanied by an enhanced sensitivity to undergo apoptosis and necrosis without inflammation. *Hepatology* 1996;**23**(2):329-37. [DOI: 1002/hep.510230220]

PG12 2019

Sullivan C, Dean J. Prescribing Guideline PG12 Pharmacological Treatment of Gender Dysphoria. Devon Partnership NHS Trust 2019.

Prior 1989

Prior CJ, Vigna YM, Watson D. Spironolactone with physiological female steroids for presurgical therapy of male-to-female transsexualism. *Archives of Sexual Behavior* 1989;**18**(1):49-57. [DOI: 18(1):49-57]

Ramsay 1990

Ramsay ID, Rushton DH. Reduced serum vitamin B12 levels during oral cyproterone-acetate and ethinyl-oestradiol therapy in women with diffuse androgen-dependent alopecia. *Clinical and Experimental Dermatology* 1990;**15**(4):277-81. [DOI: 1111/j.1365-2230.1990.tb02089.x]

Reilly 2019

Reilly ZP, Fruhauf TF, Martin SJ. Barriers to evidence-based transgender care: knowledge gaps in gender-affirming hysterectomy and oophorectomy. *Obstetrics & Gynecology* 2019;**134**(4):714-17. [DOI: 10.1097/AOG.0000000000003472]

Reisner 2016a

Reisner SL, Deutsch MB, Bhasin S, Bockting W, Brown GR, Feldman J, et al. Advancing methods for US transgender health research. *Current opinion in endocrinology, diabetes, and obesity* 2016;**23**(2):198. [DOI: 10.1097/MED.0000000000000229]

Reisner 2016b

Reisner SL, Radix A, Deutsch MB. Integrated and gender-affirming transgender clinical care and research. *Journal of acquired immune deficiency syndromes* 1999;**72**(3):235. [DOI: 10.1097/QAI.0000000000001088]

Riley 2011

Riley RD, Higgins JP, Deeks JJ. Interpretation of random effects meta-analyses. BMJ 2011;**342**. [DOI: 1136/bmj.d549.]

Schneider 2006

Schneider H, Stalla G. Hormonal Therapy [Hormonelle Therapie]. In: Therapieleitfaden Transsexualität. 1 edition. Bremen: Uni-Med Science, 2006:85-9. [ISBN: 3-89599-888-5]

Schneider 2016

Schneider C, Cerwenka S, Nieder TO, Briken P, Cohen-Kettenis PT, de Cuypere G, et al. Measuring gender dysphoria: a multicenter examination and comparison of the Utrecht gender dysphoria scale and the gender identity/gender dysphoria questionnaire for adolescents and adults. *Archives of Sexual Behavior* 2013;**45**(3):551-8. [DOI: 1007/s10508-016-0702-x]

Schulz 2010

Schulz KF, Altman DG, Moher D. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *The BMJ* 2010;**340**:698-702. [DOI: 10.1136/bmj.c332]

Schürmeyer 1986

Schürmeyer T, Graff J, Senge T, Nieschlag E. Effect of oestrogen or cyproterone acetate treatment on adrenocortical function in prostate carcinoma patients. Acta Endocrinologica 1986;**111**(3):360-7. [PMID: 2421511]

Seal 2012

Seal LJ, Franklin S, Richards C, Shishkareva A, Sinclaire C, Barret J. Predictive markers for mammoplasty and a comparison of side effect profiles in transwomen taking various hormonal regimens. *The Journal of Clinical Endocrinology & Metabolism* 2012;**97**(12):4422-8. [DOI: 1210/jc.2012-2030]

Soll 2018

Soll BM, Robles-García R, Brandelli-Costa A, Mori D, Mueller A, Vaitses-Fontanari AM, et al. Gender incongruence: a comparative study using ICD-10 and DSM-5 diagnostic criteria. *Brazilian Journal of Psychiatry* 2018;**40**(2):174-80.

Sterne 2011

Sterne JA, Sutton AJ, Ioannidis JP, Terrin N, Jones DR, Lau J, et al. Recommendations for examining and interpreting funnel plot asymmetry in meta-analyses of randomised controlled trials. BMJ 2011;**343**. [343:d4002]

Sterne 2016

Sterne JAC, Hernán MA, Reeves BC, Savović J, Berkman ND, Viswanathan M, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ* 2016;**355**:i4919. [DOI: 10.1136/bmj.i4919]

Thompson 1993

Thompson DF, Carter JR. Drug-induced gynecomastia. *Pharmacotherapy* Jan-Feb 1993;**13**(1):37-45. [PMID: 8094898]

von Elm 2007

von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, STROBE Initiative. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies.. *Annals of Internal Medicine* 2007;**147**(8):573-7. [DOI: 10.7326/0003-4819-147-8-200710160-00010] [PMID: 17938396]

Wenqing 2005

Wenqing G, Bohl CE, Dalton JT. Chemistry and structural biology of androgen receptor. *Chemical Reviews* 2005;**105**(9):3352-70. [DOI: 10.1021/cr020456u]

WHO 2018

World Health Organisation. International classification of diseases for mortality and morbidity statistics (11th Revision). icd.who.int/en/ (accessed 30 October 2020).

Wierckx 2014

Wierckx K, van Caenegem E, Schreiner T, Haraldsen I, Fisher AD, Toye K, et al. Cross-sex hormone therapy in trans persons is safe and effective at short-time follow-up: results from the European network for the investigation of gender incongruence. *International Society for Sexual Medicine* 2014;**11**(8):1999-2011. [DOI: 10.1111/jsm.12571]

WPATH 2011

Coleman E, Bockting W, Botzer M, Cohen-Kettenis P, DeCuypere G, Feldman J, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people version 7. *International Journal of Transgenderism* 2011;**13**(4):165–232. [DOI: 10.1080/15532739.2011.700873]

Zucker 2016

Zucker KJ. The DSM-5 diagnostic criteria for gender dysphoria. In: Trombetta C, Liguori G, Bertolotto M, editors(s). Management of Gender Dysphoria - A Multidisciplinary Approach. First edition. Vol. **1**. Mailand: Springer, 2016:33-7. [ISBN: 978-88-470-5695-4]

References to other published versions of this review

Haupt 2018

Haupt C, Henke M, Kutschmar A, Hauser B, Baldinger S, Schreiber G. Antiandrogens or estradiol treatments or both during hormone replacement therapy in transitioning transgender women. *Cochrane Database of Systematic Reviews* 2018, Issue 10. Art. No: CD013138. [DOI: 10.1002/14651858.CD013138]

CHARACTERISTICS OF STUDIES

Characteristics of excluded studies *[ordered by study ID]*

Study	Reason for exclusion
Asscheman 2011	Mortality rates in transgender people receiving long-term cross-sex hormones. A cohort study. Adequate controls are missing. Interventions are not clearly defined
Colizzi 2015	Increased prevalence of metabolic syndrome among individuals with gender dysphoria treated by cross-sex hormonal treatment. Study without adequate comparator group.
Fighera 2018	Hormone therapy has been associated with changes in bone and lean/fat mass. This study assessed bone mineral density, appendicular lean mass, and total fat mass in transwomen undergoing cross-sex hormone therapy. Study without adequate comparator group.
Fisher 2014	This study aimed to assess differences in body uneasiness and psychiatric symptoms between gender dysphoria clients taking hormone therapy and those not taking hormones (no hormone therapy). A second aim was to assess whether length of hormone treatment and daily dose provided an explanation for levels of body uneasiness and psychiatric symptoms. Cross-sectional design.
Fisher 2016	The objective of the study was to assess whether hormone therapy-related body changes affect psychobiological well-being in gender dysphoria. Study without adequate comparator group.
Giltay 2000	Hormone therapy effects on the skin (hair growth rate, density, and shaft diameter by image analysis; and sebum production) of transsexual patients receiving cross-sex hormones. It is a case series, adequate controls are missing.
Haraldsen 2005	Hormone therapy effects on cognitive performance. Study without adequate comparator group.
Haraldsen 2007	The effects of cross-sex hormones on bone metabolism (bone mineral density, total body fat, total lean body mass) in patients with early onset gender identity disorder. Study without adequate comparator group.
Miles 2006	The study was designed to examine associations between oestrogen and cognition (memory, including visual, spatial, object and location memory, other cognitive abilities that show reliable sex

Study	Reason for exclusion
	differences, including verbal and visual-spatial abilities, and mood variables). The cross-over design used was comparative, but did not used randomization or quasi-randomisation.
Schlatterer 1998	This follow-up study was carried out to validate the effectiveness of cross-gender hormone therapy embedded in a multistep treatment concept for transgender patients. Study without adequate comparator group. This study lacks adequate controls.
Toorians 2003	To find an explanation for the different thrombotic risks of oral ethinyl estradiol and transdermal 17-beta-estradiol use, the researchers compared the effects of treatment of male-to-female transgender people with cyproterone acetate only, and with cyproterone acetate in combination with transdermal 17-beta-estradiol, oral ethinyl estradiol, or oral 17-beta-estradiol on a number of haemostatic variables. There is no adequate control group.
Van Goozen 1995	Effects of sex hormones to the establishment of gender differences in behaviour, a large battery of tests on aggression, sexual motivation and cognitive functioning was administered twice: shortly before and three months after the start of cross-sex hormone treatment. The study does not have an adequate comparator group.

Characteristics of ongoing studies *[ordered by study ID]*

Krasean 2019

Study name	Anti-androgenic effects comparison between cyproterone acetate and spironolactone in transgender women: a randomised controlled trial (Trial ID: TCTR20190404001)
Methods	Allocation: randomised Study design: randomised controlled trial Control: active Study endpoint classification: efficacy study Intervention model: Parallel Number of arms: 2 Masking: double blind (Masked roles: participant caregiver, investigator) Primary purpose: treatment Study phase: phase 4
Participants	Gender: male Age limit: minimum 18 years: maximum 40 years Condition: Gender dysphoria patients diagnosed from DSM V Male to female transgender Not undergone orchidectomy No psychological disease or mental disability
Interventions	Arm 1: Intervention name: cyproterone acetate Type: active comparator

Krasean 2019 *(Continued)*

	Classification: drug Descriptions: participants (gender dysphoria patients) will receive estradiol valerate (4 mg daily) combined with cyproterone acetate (25 mg daily) for cross-sex hormone treatment. Arm: 2 Intervention name: spironolactone Type: experimental Classification: drug Descriptions: participants (gender dysphoria patients) will be received estradiol valerate (4 mg daily) combined with spironolactone (100 mg daily) for cross-sex hormone treatment.
Outcomes	Primary outcome(s): Outcome name: testosterone level Measurement: Electrochemiluminescent Immunoassay (ECLIA) of total testosterone level Time point: three months after intervention Safety issue: no Key secondary outcomes: Outcome name: physical and metabolic changes Measurement: physical examination, metabolic profile parameters Time point: three months after intervention Safety Issue: no
Starting date	April 3, 2019 (estimated end date: June 16, 2020)
Contact information	Contact: Krasean Panyakhamlerd Degree: Assoc. Prof. Phone: 0926536415 Email: krasean@hotmail.com Postal Address: 1873 Rama 4 Road, Patumwan State/Province: Bangkok Postal Code: 10400 Country: Thailand
Notes	Source(s) of monetary or material supports: Ratchadapisek Sompoch Fund, Faculty of Medicine, Chulalongkorn University Declarations of interest not reported

ADDITIONAL TABLES

Table 1. Suggested design of future studies

Methods	RCT or controlled cohort study
Participants	Transgender women experiencing gender dysphoria, in transition N* Age: from the age of 16 years
Intervention	• Antiandrogens (cyproterone acetate or spironolactone) and estradiol • Antiandrogens (cyproterone acetate or spironolactone) alone • Estradiol alone All types of administration: oral, sublingual, transdermal, subdermal and intramuscular. For estradiol and bioidentical 17-beta-estradiol, as well as synthetic derivatives.
Comparator	Any of the active interventions listed above
Outcomes	Primary outcomes • Quality of life (QoL) • Satisfaction with change of male to female body characteristics, • Adverse events specific to hormone therapy, including serious adverse events
Notes	* Size of study with sufficient power to detect a ~ 10% difference between the two groups for primary outcome

APPENDICES

Appendix 1. OvidSP search strategy

Search	Query
#1	(transsexual* OR transgender OR "gender dysphoria" OR transident* OR "trans women" OR "trans woman").mp.
#2	("cyproterone acetate" OR CPA OR androcur).mp. or cyproterone Acetate/
#3	(spironolactone OR Aldactone OR Jenaspiron OR Osyrol OR Spirobene OR Verospiron OR Xenalon).mp. or spironolactone/
#4	(estradiol* OR oestradiol* OR estrifam OR gynocadin OR neofollin OR lenzetto).mp. or Estradiol/
#5	2 OR 3 OR 4
#6	1 AND 5

HISTORY

Protocol first published: Issue 10, 2018
Review first published: Issue 11, 2020

CONTRIBUTIONS OF AUTHORS

All authors contributed to the Abstract, Background, Methods, Results, Discussion, and Authors' conclusions.
Claudia Haupt, Alexia Kutschmar and Miriam Henke conducted the study selection.

DECLARATIONS OF INTEREST

Claudia Haupt declares no competing interest.

Miriam Henke declares no competing interest.

Alexia Kutschmar declares no competing interest.

Birgit Hauser (BH) declares no competing interest. BH is a clinical practitioner in private practice, who also prescribes hormone therapy.

Sandra Baldinger declares no competing interest.

Sarah Rafaela Saenz declares no competing interest.

Gerhard Schreiber declares no competing interest.

None of the review authors' incomes depends on the prescription of drugs. The review authors did not receive any financial support for this project, but paid for all related expenses themselves. They worked voluntarily and free of charge.

INDEX TERMS

Medical Subject Headings (MeSH)

Androgen Antagonists [*therapeutic use]; Drug Therapy, Combination [methods]; Estradiol [*therapeutic use]; Estrogens [*therapeutic use]; Placebos [therapeutic use]; Sex Reassignment Procedures [*methods]; *Transgender Persons

MeSH check words

Female; Humans; Male

Acta Psychiatrica Scandinavica

Acta Psychiatr Scand 2020: 141: 486–491

DOI: 10.1111/acps.13164

ACTA PSYCHIATRICA SCANDINAVICA

Trends in suicide death risk in transgender people: results from the Amsterdam Cohort of Gender Dysphoria study (1972–2017)

Wiepjes CM, den Heijer M, Bremmer MA, Nota NM, de Blok CJM, Coumou BJG, Steensma TD. Trends in suicide death risk in transgender people: results from the Amsterdam Cohort of Gender Dysphoria study (1972–2017).

Objective: This study explored the overall suicide death rate, the incidence over time, and the stage in transition where suicide deaths were observed in transgender people.
Methods: A chart study, including all 8263 referrals to our clinic since 1972. Information on death occurrence, time, and cause of death was obtained from multiple sources.
Results: Out of 5107 trans women (median age at first visit 28 years, median follow-up time 10 years) and 3156 trans men (median age at first visit 20 years, median follow-up time 5 years), 41 trans women and 8 trans men died by suicide. In trans women, suicide deaths decreased over time, while it did not change in trans men. Of all suicide deaths, 14 people were no longer in treatment, 35 were in treatment in the previous two years. The mean number of suicides in the years 2013–2017 was higher in the trans population compared with the Dutch population.
Conclusions: We observed no increase in suicide death risk over time and even a decrease in suicide death risk in trans women. However, the suicide risk in transgender people is higher than in the general population and seems to occur during every stage of transitioning. It is important to have specific attention for suicide risk in the counseling of this population and in providing suicide prevention programs.

C. M. Wiepjes[1,2], M. den Heijer[1,2], M. A. Bremmer[2,3], N. M. Nota[1,2], C. J. M. de Blok[1,2], B. J. G. Coumou[2,4], T. D. Steensma[2,4]

[1]Department of Endocrinology, Amsterdam UMC, VU University Medical Center, Amsterdam, the Netherlands, [2]Center of Expertise on Gender Dysphoria, Amsterdam UMC, VU University Medical Center, Amsterdam, the Netherlands, [3]Department of Psychiatry, Amsterdam UMC, VU University Medical Center, Amsterdam, the Netherlands and [4]Department of Medical Psychology, Amsterdam UMC, VU University Medical Center, Amsterdam, the Netherlands

Key words: gender dysphoria; transgender; suicide

Thomas D. Steensma, Department of Medical Psychology, Amsterdam UMC, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, the Netherlands. E-mail: t.steensma@amsterdamumc.nl

Accepted for publication February 16, 2020

Significant outcomes

- Suicide death risk in trans people did not increase over time.
- Suicide deaths occurred during every stage of transitioning.
- Suicide death risk is higher in trans people than in the general population.

Limitations

- Psychological comorbidity was not known.
- No data were available for people on the waiting list for their first appointment.

Introduction

Gender dysphoria (GD) refers to the distress related to a marked incongruence between one's assigned gender at birth and the experienced gender (1). Trans people are diverse in the intensity of experienced GD (2), their needs for medical

transition (3), and the impairment that GD can have on their life. Studies focusing on the wellbeing of trans people show a greater vulnerability for experiencing mental health problems compared with the non-trans (cis) population (4). Most prevalent are affective and anxiety problems (5-7), often accompanied by feelings, thoughts, or behaviours linked to suicidality (8,9).

The prevalence of suicidality in trans people in suicidal ideation, suicidal attempts, and suicide death rates is studied in varying degrees and shows high variability in findings. A systematic review by McNeil et al (9). reported suicidal ideation rates across 17 identified studies, ranging from 37% (10) up to 83% (11). Prevalence rates on suicidal attempts in trans people, which are generally observed to be lower than suicidal ideation, showed to be lower but also with a wide variation in reported rates, ranging from 9.8% (12) up to 44% (13). Since structured prevalence studies on suicide deaths are lacking in the transgender literature, an estimation comes from a limited number of studies reporting on suicide death rates in small study samples. Derived from a systematic review on suicidality in trans people by Marshall et al. (8), suicide death rates varied from 0% (14) to 4.2% in a sample of 24 post-treatment trans people from Sweden (15). Six of these studies only included postsurgical people (14-19), whereas two studies also included trans people who were only using hormones without surgery (20,21). However, studies differentiating the treatment stage during which death by suicide occurred are lacking. In addition, studies differentiating between suicide in trans women and trans men are scarce. While some studies found that trans men have a higher risk of suicide attempts than trans women (22,23), other studies reported no differences in suicide attempts between trans women and trans men (24,25). Only one cohort distinguished suicide death risk in trans women and trans men and found that trans women had an increased risk of suicide death compared with trans men (20,21).

Aims of the study

The aim of the current study is to explore the overall suicide death rate in trans women and trans men in the largest clinical cohort of gender-referred people seen at the Center of Expertise on Gender Dysphoria of the Amsterdam University Medical Centers between 1972 and 2017 the Netherlands (26). In addition, the change in incidence of suicide death rate over time and at what stage in transition (pretreatment, during hormonal treatment and/or surgical phase, or post-treatment) suicide deaths were observed was explored. The relevance of such information is to get a greater understanding of how large the risk is in clinically referred transgender people and whether suicide prevention interventions should focus on specific stages in transition or not.

Material and methods

Study design

A retrospective chart study was performed, including all people who once visited the Center of Expertise on Gender Dysphoria of the Amsterdam UMC, Vrije Universiteit Amsterdam, the Netherlands, between 1972 and 2017. The selection of the study population is described previously (26). A total of 8263 adults, adolescents, and children were included, with a median age at first visit of 25 years (range 4 to 81 years) and a median follow-up time of 7.5 years (range 0.0 to 45.5 years). Information on death occurrence, time, and cause of death was obtained by cross-checking multiple sources: the National Civil Record Registry (21), which contains date of birth and date of death of all inhabitants of the Netherlands, and the hospital registration system, medical, and psychological files for cause of death.

The Medical Ethics Review Committee of the Amsterdam UMC, Vrije Universiteit Amsterdam, reviewed this study and determined that the Medical Research Involving Human Subjects Act (WMO) did not apply to this study. Therefore, and because of the retrospective design, necessity for informed consent was waived.

Treatment

After an initial visit to the endocrinologist (for adults) or child psychiatrist (for children and adolescents), all people were referred to the psychology department for the diagnostic phase. In this phase, people were seen to gain insight into their experienced gender identity, to verify whether they fulfill the diagnosis gender dysphoria, to explore their treatment desires, and to prepare them for possible medical interventions. After this phase, people may start with hormonal treatment. Trans women received treatment with anti-androgens and estrogens. Trans men were treated with testosterone. In adolescents, treatment first started with a period of puberty suppression, followed by estrogens of testosterone around the age of 16 years (27).

Surgical interventions can be offered to people aged 18 years or older. Depending on the desired

treatment, the surgery is preceded after at least one year of hormonal treatment (genital surgery) or can be offered after the diagnostic phase (e.g., breast removal). After surgery, all people were usually seen every 2 years for medical check-up.

Statistical analyses

Characteristics of the population were shown as median with range due to the non-normal distribution. The total number of people seen at our center and the total number of suicide deaths were counted and were expressed as percentages as well as incidence per 100 000 person years. For each year, the number of people at risk and the number of people who died by suicide were calculated. Cox regression analyses were performed to calculate hazard ratios (HR) with corresponding 95% confidence intervals (95% CI). Date of first visit was used as start date of follow-up. The end date of follow-up was either date of death or date of closing the database (December 31, 2017). Suicide death was analyzed as event. To analyze whether the incidence of suicide deaths changed over time, the year of first visit was added as determinant to the analyses. Analyses were adjusted for age at first visit as age might be related to suicide death risk. Time between date of suicide death and first visit, and between date of suicide death and start of hormonal treatment, if applicable, were calculated. All analyses were performed for the total population and were stratified for trans women and trans men.

All analyses were performed using STATA Statistical software (Statacorp, College Station, TX, USA), version 15.1.

Results

The characteristics of the study population are shown in Table 1. In total, 8263 people attended the gender identity clinic, of which 5107 were trans women (median age at first visit 28 years, range 4 to 81 years) and 3156 were trans men (median age at first visit 20 years, range 4 to 73 years). The median follow-up time was 7.5 years (range 0.0–45.5 years), which was longer in trans women (10.2 years, range 0.0–45.5 years) than in trans men (4.8 years, range 0.0–45.5 years). The total follow-up time was 92 227 person years (64 287 in trans women and 27 940 in trans men).

Table 1. Characteristics of the study population (A) and the people who died by suicide (B)

	Total	Trans women	Trans men
(A)			
Number of people	8263	5107	3156
Age at first visit, year	25 (4–81)	28 (4–81)	20 (4–73)
Follow-up time, year	7.5 (0.0–45.5)	10.2 (0.0–45.5)	4.8 (0.0–45.5)
(B)			
Number of suicides	49 (0.6%)	41 (0.8%)	8 (0.3%)
Age at first visit, year	31 (15–59)	31 (15–58)	21 (16–59)
Age at time of suicide, year	41 (18–66)	41 (18–66)	36 (21–60)
Follow-up time, year	6.7 (0.6–32.7)	6.7 (0.6–32.7)	6.7 (0.6–23.1)
Time between start	6.4 (0.4–32.5)	6.1 (0.4–32.5)	6.9 (3.7–23.1)
hormones and suicide, year	*n = 42*	*n = 35*	*n = 7*

Data are shown as number or median (range).

Forty-nine people died by suicide: 41 trans women (0.8%) and 8 trans men (0.3%), which is 64 per 100 000 person years in trans women and 29 per 100 000 person years in trans men. The median follow-up time between first visit and suicide death was 6.7 years (range 0.6 to 32.7 years) in trans women and 6.7 years (range 0.6 to 23.1 years) in trans men. Trans women had a higher overall suicide death risk than trans men (per year: HR 2.26, 95% CI 1.06–4.82). Four suicide deaths occurred in individuals who were referred to the clinic before the age of 18 (0.2%), which is a lower risk than in adults (0.7%, $P = 0.010$).

The course of number of people at risk and the number of people who died by suicide over the years is shown in Fig. 1. Overall suicide deaths did not increase over the years: HR per year 0.97 (95% CI 0.94–1.00). In trans women, suicide death rates decreased slightly over time (per year: HR 0.96, 95% CI 0.93–0.99), while it did not change in trans men (per year: HR 1.10, 95% CI 0.97–1.25). Adjustment for age at the first visit did not change these numbers.

As the median follow-up time between first visit and suicide death was 6.7 years, subgroup analyses were performed in those who had their first visit before 2011. This did not change the outcomes: trans women ($n = 3115$) HR 0.94, 95% CI 0.91–0.98; trans men ($n = 1269$) HR 1.02, 95% CI 0.90–1.16).

Of the 49 people who died by suicide, 35 had a face-to-face contact with the endocrinologist or psychologist of the gender identity clinic in the previous two years, while the other 14 people were no longer in active counseling with the clinic. Sixteen of the 35 people who recently had visited the clinic, only came for a medical check-up, as they were postsurgery (vaginoplasty or phalloplasty). Two people were in the surgery trajectory, and 17 were still in the diagnostic or hormonal phase at time of suicide. The transition phases separately for trans women and trans men who died by suicide are shown in Table 2.

The mean number of suicides in the years 2013–2017 was higher in the trans population (40 per 100 000 person years; 43 per 100 000 trans women

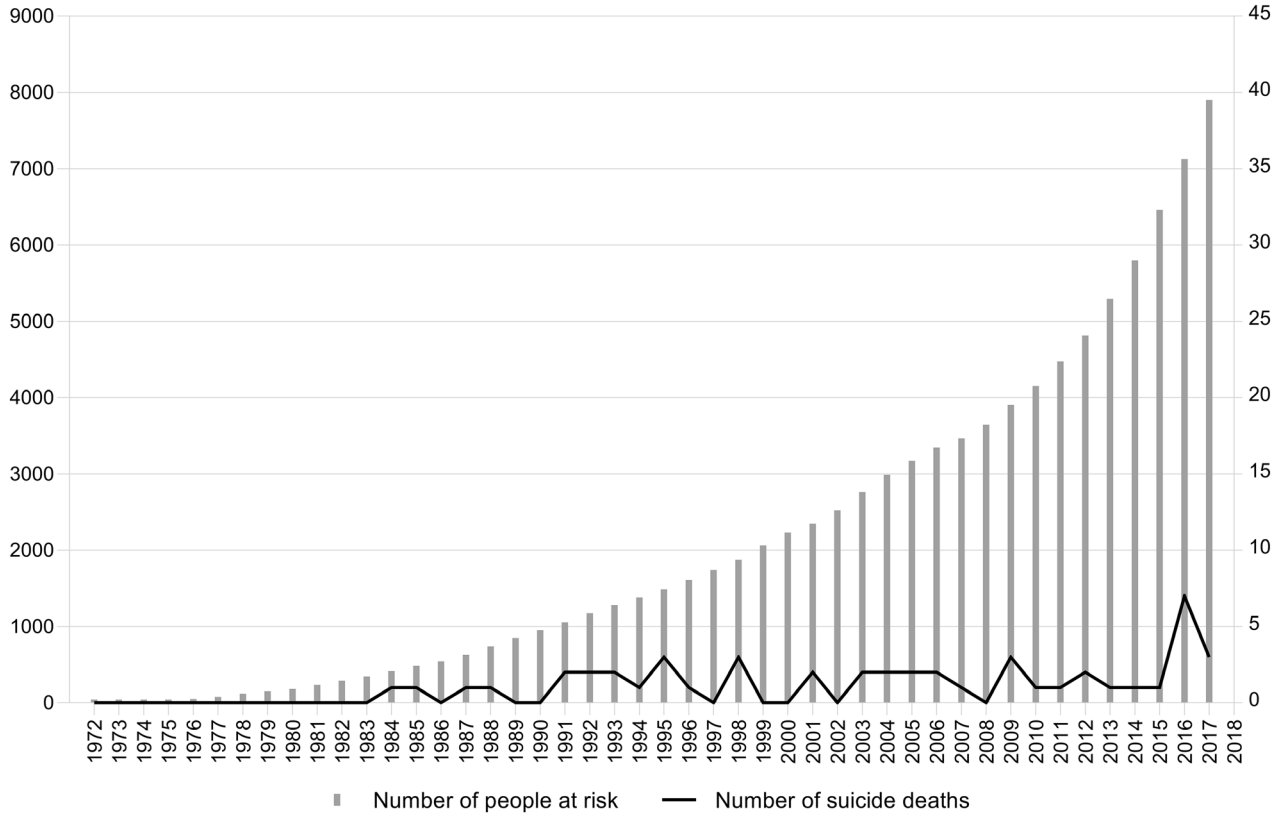


Fig. 1. Number of people at risk (left *y*-axis) and the number of suicides (right *y*-axis), between 1972 and 2017.

Table 2. The occurrence of suicide deaths distinguished for transition stage, and trans women or trans men

	Total (n = 49)	Trans women (n = 41)	Trans men (n = 8)
In active counseling	35	29	6
In diagnostic or hormonal phase	17	16	1
In surgical phase	2	0	2
Only medical follow-up care	16	13	3
No active counseling	14	12	2

Data are shown as number. In active counseling is defined as a face-to-face contact with the endocrinologist or psychologist of the gender identity clinic in the previous two years.

and 34 per 100 000 trans men) compared with the Dutch population in this time frame (11 per 100 000 person years; 15 per 100 00 registered men and 7 per 100 000 registered women) (28).

Discussion

The current study investigated the suicide death risk in the largest clinical cohort of gender-referred individuals to the Center of Expertise on Gender Dysphoria at the Amsterdam UMC, the Netherlands, between 1972 and 2017. Findings from the chart reviews showed us a decrease in suicide death risk over time in trans women and no change in suicide death risk in trans men. Trans women, however, showed a higher suicide death risk than trans men. Between 2013 and 2017, the suicide risk in Dutch referred transgender people (40 per 100 000 person years) showed to be three to four times higher than the general Dutch population (11 per 100 000 person years) (28). Evaluation of transition stage in relation to suicide deaths showed that approximately two-third of the observed suicides occurred in those who were still in active treatment (diagnostic, hormonal, or surgical phase). The incidence of suicide deaths and transition stage was similar in trans women and trans men.

Suicidal behaviour is a complex phenomenon that is a result of many individual (age, male sex assigned at birth, previous suicide attempts, mental health history, substance abuse) as well as more distant environmental factors. A recent literature review clearly demonstrates the specific risk factors for suicide in sexual minority youth, which includes negative social environments, inadequate support within the closest social network, and an absence of lesbian, gay, bisexual, and transgender (LGBT) movements in communities (29). In our cohort, both trans women and trans men show a three- to four-fold elevated

risk of suicide compared with the population rate in the Netherlands and can therefore be considered a high-risk group. Although the Netherlands is known for its tolerance toward sexual minority groups in comparison to most countries in the world (30), the societal position of trans people is generally less favorable compared with the lesbian, gay, bisexual, and cisgender population. Furthermore, compared with trans men, the societal position of trans women is lower (31,32).

In the Netherlands, between 1972 and 2017 suicide rates showed a fluctuating course. Our finding of a slightly decreasing suicide risk in Dutch trans women may confer some hope. Recent studies showed an increase in societal acceptance toward lesbian, gay, bisexual, and transgender people (31), and indications of an increase in social-economic status over the years (33). Although specific information on trans men and trans women is unavailable, it is conceivable that the improvement of societal position may have effect on the psychological functioning and the prevention of suicidal risk in trans women. The cause of this increase in tolerance seems largely to be the effect of a national and international increase in visibility and attention for trans people in media and society. Another explanation may be that, with the increase in attention and acceptance, the threshold for transgender people to seek treatment or professional help has become lower over the years. This is also reflected by the increase in referrals each year (26). Lastly, with the increase of knowledge in this field and the literature about the vulnerability of the transgender population for suicidal ideation, suicidal attempts, and suicide death rates, it is conceivable to assume that the attention to these risks has increased in clinical counseling and may have its effect on prevention of suicide deaths over the years.

Although the incidence of suicide deaths in trans women decreased over the years, the overall incidence still showed to be higher in trans women compared with trans men. Conflicting results in literature are reported about the risk of suicide attempts between trans women and trans men. Some studies reported that trans men had a higher risk of suicide attempts than trans women (22,23), while in other studies no differences in suicide attempts between trans women and trans men were found (24,25). Only two studies looked at the differences in the risk of death by suicide between trans women and trans men and found that trans women had an increased risk compared with trans men (20,21). However, these two studies were earlier studies performed in our center and therefore include a smaller part of our current study population.

An important finding was that the incidence for observed suicide deaths was almost equally distributed over the different stages of treatment. Although the distribution showed that one-third of the suicides occurred in people who were no longer in active treatment in our center, the other two-third of the people who died by suicide still visited our center in the previous two years. About half of these last two-third people were still in active diagnostic or medical treatment, while the other half completed their transition and only came for a medical check-up. This indicates that vulnerability for suicide occurs similarly in the different stages of transition. Although the literature on suicide risk factors is comprehensive, and particular suicidal risk factors like verbal victimization, physical and sexual violence, and the absence of social support (9,34), may apply for transgender people in all stages of transitioning, it seems clinically highly relevant to understand and explore possible differences in motives and risk factors in the different stages of treatment. Therefore, future research on suicide deaths and suicide risk factors in transgender people should have a greater focus on transition status in relation to these motives and risk factors.

This study is performed in the largest cohort of gender-referred people from the Netherlands, consisting of a large population of both adult and adolescent trans women and trans men at different stages of their transition with a long follow-up time. However, this study has also some limitations. First, this study is a retrospective chart study. Although we used multiple strategies to obtain data about date of death, it is possible that we missed some data. Second, we did not have information about psychological comorbidities or other psychological information, such as social support. Third, we only had information about people who actually visited our gender identity clinic. Information about people on the waiting list for their first appointment was lacking.

To conclude, in our clinic we observed no increase in suicide death risk over time and even a decrease over time in suicide death risk in trans women was found. Since the suicide risk in the transgender population is higher than the general population and seems to occur during every stage of transitioning, it is important that (mental) health practitioners pay attention to this risk and create a safe environment in which these feelings can be discussed at all stages of treatment and counseling. Further research is necessary to investigate the motives behind the suicides, as input in

the development of adequate suicide prevention programs.

Conflicts of interests

None.

Data availability statement

Author elects to not share data.

References

1. American Psychiatry Association. Diagnostic and Statistical Manual of Mental Disorders (DSM-5), 5th ed. Arlington, VA: American Psychiatric Publishing; 2013.
2. Meyer-Bahlburg HF. From mental disorder to iatrogenic hypogonadism: dilemmas in conceptualizing gender identity variants as psychiatric conditions. Arch Sex Behav 2010;**39**(2):461–476.
3. Beek TF, Kreukels BP, Cohen-Kettenis PT, Steensma TD. Partial treatment requests and underlying motives of applicants for gender affirming interventions. J Sex Med 2015;**12**(11):2201–2205.
4. Dhejne C, Van Vlerken R, Heylens G, Arcelus J. Mental health and gender dysphoria: A review of the literature. Int Rev Psychiatry 2016;**28**(1):44–57.
5. Haraldsen IR, Dahl AA. Symptom profiles of gender dysphoric patients of transsexual type compared with patients with personality disorders and healthy adults. Acta Psychiatr Scand 2000;**102**:276–281.
6. Fisher AD, Bandini E, Casale H et al. Sociodemographic and clinical features of gender identity disorder: an Italian multicentric evaluation. J Sex Med 2013;**10**(2):408–419.
7. Heylens G, Elaut E, Kreukels BP et al. Psychiatric characteristics in transsexual individuals: multicentre study in four European countries. Br J Psychiatry 2014;**204**(2):151–156.
8. Marshall E, Claes L, Bouman WP, Witcomb GL, Arcelus J. Non-suicidal self-injury and suicidality in trans people: A systematic review of the literature. Int Rev Psychiatry 2016;**28**(1):58–69.
9. McNeil J, Ellis SJ, Eccles FJR. Suicide in trans populations: A systematic review of prevalence and correlates. Psychol Sex Orient Gender Div 2017;**4**(3):341–353.
10. Mathy RM. Transgender Identity and Suicidality in a Nonclinical Sample. J Psychol Human Sexual. 2003;**14**(4):47–65.
11. Testa RJ, Sciacca LM, Wang F et al. Effects of Violence on Transgender People. Profess Psychol: Res Pract 2012;**43**(5):452–459.
12. Heylens G, Verroken C, De Cock S, T'Sjoen G, De Cuypere G. Effects of different steps in gender reassignment therapy on psychopathology: a prospective study of persons with a gender identity disorder. J Sex Med 2014;**11**(1):119–126.
13. Miller LR, Grollman EA. The social costs of gender nonconformity for transgender adults: implications for discrimination and health. Sociol Forum 2015;**30**(3):809–831.
14. Johansson A, Sundbom E, Hojerback T, Bodlund O. A five-year follow-up study of Swedish adults with gender identity disorder. Archiv Sex Behav 2010;**39**(6):1429–1437.
15. Walinder J, Law ITA. Concerning Sex Reassignment of Transsexuals in Sweden. Archiv Sex Behav 1976;**5**(3):255–258.
16. Dhejne C, Lichtenstein P, Boman M, Johansson AL, Langstrom N, Landen M. Long-term follow-up of transsexual persons undergoing sex reassignment surgery: cohort study in Sweden. PLoS One 2011;**6**(2):e16885.
17. Berg JEA, Gustafsson M. Long term follow up after sex reassignment surgery. Scand J Plast Reconstruct Surg Hand Surg 2009;**31**(1):39–45.
18. De Cuypere G, T'Sjoen G, Beerten R et al. Sexual and physical health after sex reassignment surgery. Arch Sex Behav. 2005;**34**(6):679–690.
19. Sorensen T, Hertoft P. Male and female transsexualism: the danish experience with 37 patients. Archiv Sex Behav 1982;**11**(2):133–155.
20. Van Kesteren P, Asscheman H, Megens JAJ, Gooren LJG. Mortality and morbidity in transsexual subjects treated with cross-sex hormones. Clin Endocrinol 1997;**47**:337–342.
21. Asscheman H, Giltay EJ, Megens JA, de Ronde WP, van Trotsenburg MA, Gooren LJ. A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones. Eur J Endocrinol 2011;**164**(4):635–642.
22. Marshall BD, Socias ME, Kerr T, Zalazar V, Sued O, Aristegui I. Prevalence and correlates of lifetime suicide attempts among transgender persons in Argentina. J Homosex 2016;**63**(7):955–967.
23. Maguen S, Shipherd JC. Suicide risk among transgender individuals. Psychol Sexual 2010;**1**(1):34–43.
24. Reisner SL, White JM, Bradford JB, Mimiaga MJ. Transgender Health disparities: COMPARING full cohort and nested matched-pair study designs in a community health center. LGBT Health 2014;**1**(3):177–184.
25. Skagerberg E, Parkinson R, Carmichael P. Self-harming thoughts and behaviors in a group of children and adolescents with gender dysphoria. Int J Transgend 2013;**14**(2):86–92.
26. Wiepjes CM, Nota NM, de Blok CJM et al. The amsterdam cohort of gender Dysphoria study (1972–2015): Trends in prevalence, treatment, and regrets. J Sex Med 2018;**15**(4):582–590.
27. Kreukels BP, Cohen-Kettenis PT. Puberty suppression in gender identity disorder: the Amsterdam experience. Nat Rev Endocrinol 2011;**7**(8):466–472.
28. StatLine CBSOverledenen; belangrijke doodsoorzaken (korte lijst), leeftijd, geslacht; 2018.
29. Poštuvan V, Podlogar T, Zadravec Šedivy N, De Leo D. Suicidal behaviour among sexual-minority youth: a review of the role of acceptance and support. Lancet Child Adol Health 2019;**3**(3):190–198.
30. Veenhoven R. Wat bracht de seksuele revolutie (What caused the sexual revolution?). In: Couwenberg SW ed. Seksuele Revolutie Ter Discussie. Van Phil Bloom tot Sex and the Citiy. Vol 1. Eindhoven, the Netherlands: Damon, 2005:93-105.
31. Kuyper L. Opvattingen over seksuele en genderdiversiteit in Nederland en Europa. Den Haag: Sociaal en Cultureel Planbureau, 2018.
32. Keuzenkamp S. Worden wie je bent. Het leven van transgenders in Nederland. Den Haag: Sociaal en Cultureel Planbureau, 2012.
33. Beusekom G, Kuyper L. LHBT-monitor 2018. De leefsituatie van lesbische, homoseksuele, biseksuele en transgender personen in Nederland. Den Haag: Sociaal en Cultureel Planbureau, 2018.
34. Haas AP, Rodgers PL, Herman JL. Suicide attempts among transgender and gender non-conforming adults. findings of the national transgender discrimination survey. New York, NY: American Foundation for Suicide Prevention and The Williams Institute, 2014.

Socialstyrelsen
THE NATIONAL BOARD OF HEALTH AND WELFARE

Care of children and adolescents with gender dysphoria

Summary

Summary

The National Board of Health and Welfare (NBHW) has been commissioned by the Swedish government to update the national guidelines on care of children and adolescents with gender dysphoria, first published in 2015 [1]. Guidelines chapters are updated stepwise and this report contains revised guidance on psychosocial support and diagnostic assessment, and on puberty suppressing treatment with GnRH-analogues and gender-affirming hormonal treatment. This report thus replaces the corresponding chapters in the publication from 2015. Remaining chapters and the updated guidelines as a whole will be published later in 2022. In response to comments received during external review, two new chapters have been added, named *New recommendations on hormonal treatment – their reasons and consequences and Non-binary gender identity – current knowledge and a need for clarification*. Another difference compared to the guidelines from 2015 [1] is that the term "gender incongruence" is used alongside the term "gender dysphoria". For explanations of terms and abbreviations, see Appendix 2. For a description of the scientific evidence and clinical experience underlying the recommendations and the work process, see Appendices 3 and 4.

The guidelines apply to children and adolescents, i.e. people under 18 years of age. In the medical text sections, the term children (barn) refers to persons who have not yet entered puberty, while the term adolescents (ungdomar) refers to people whose puberty has started. In the text sections relating to juridical regulations, only the term children (barn) is used and denotes people younger than 18 years of age. Finally, the term "young people" (unga) is sometimes used in text sections addressing both children and adolescents.

Introductory comment

The summary that follows and the introductory chapter describe that the updated recommendations for puberty suppression with GnRH-analogues and gender-affirming hormonal treatment have become more restrictive compared to 2015, and the reasons that they have changed. The new recommendations entail that a larger

proportion than before, among adolescents with gender incongruence referred for diagnostic assessment of gender dysphoria, will need to be offered other care than hormonal treatments. Questions on how to ensure that all young people suffering from gender dysphoria be taken seriously and confirmed in their gender identity, well received and offered adequate care are becoming increasingly relevant, and will need to be answered during the ongoing restructuring of certain care for gender dysphoria into three national specialised medical care services (NBHW decision in December 2020). The care for children, adolescents and adults with gender dysphoria in these three national specialised units aims to improve equality in care, coordination and dialogue, and may enhance the implementation of national guidelines.

Recommendations and criteria for hormonal treatment

For adolescents with gender incongruence, the NBHW deems that the risks of puberty suppressing treatment with GnRH-analogues and gender-affirming hormonal treatment currently outweigh the possible benefits, and that the treatments should be offered only in exceptional cases. This judgement is based mainly on three factors: the continued lack of reliable scientific evidence concerning the efficacy and the safety of both treatments [2], the new knowledge that detransition occurs among young adults [3], and the uncertainty that follows from the yet unexplained increase in the number of care seekers, an increase particularly large among adolescents registered as females at birth [4].

A systematic review published in 2022 by the Swedish Agency for Health Technology Assessment and Assessment of Social Services [2] shows that the state of knowledge largely remains unchanged compared to 2015. High quality trials such as RCTs are still lacking and the evidence on treatment efficacy and safety is still insufficient and inconclusive for all reported outcomes. Further, it is not possible to determine how common it is for adolescents who undergo gender-affirming treatment to later change their perception of their gender identity or interrupt an ongoing treatment. An important difference compared to 2015 however, is that the occurrence of

detransition among young adults is now documented [3], meaning that the uncertain evidence that indicates a low prevalence of treatment interruptions or any aspects of regret is no longer unchallenged. Although the prevalence of detransition is still unknown, the knowledge that it occurs and that genderconfirming treatment thus may lead to a deteriorating of health and quality of life (i.e. harm), is important for the overall judgement and recommendation.

To minimize the risk that a young person with gender incongruence later will regret a gender-affirming treatment, the NBHW deems that the criteria for offering GnRH-analogue and gender-affirming hormones should link more closely to those used in the Dutch protocol, where the duration of gender incongruence over time is emphasized [5-7]. Accordingly, an early (childhood) onset of gender incongruence, persistence of gender incongruence until puberty and a marked psychological strain in response to pubertal development is among the recommended criteria. The publications that describe these criteria and the treatment outcomes when given in accordance [5, 6, 8] consitute the best available knowledge and should be used as guidance.

To ensure that new knowledge is gathered, the NBHW further deems that treatment with GnRH-analogues and sex hormones for young people should be provided within a research context, which does not necessarily imply the use of randomized controlled trials (RCTs). As in other healthcare areas where it is difficult to conduct RCTs while retaining sufficient internal validity, it is also important that other prospective study designs are considered for ethical review and that register studies are made possible. Until a research study is in place, the NBHW deems that treatment with GnRH-analogues and sex hormones may be given in exceptional cases, in accordance with the updated recommendations and criteria described in the guidelines. The complex multidisciplinary assessments will eventually be carried out in the three national units that are granted permission to provide highly specialized care services.

In accordance with the DSM-5, the recommendations in the guidelines from 2015 applied to young people with gender dysphoria in general, i.e. also young people with a non-binary gender identity. Another criterion within the Dutch protocol is that the child has had a binary ("cross-gender") gender identity since childhood [5, 6].

It has emerged during the review process, that the clinical experience and documentation of puberty-suppressing and hormonal treatment for young people with non-binary gender identity is lacking, and also that it is limited for adults. The NBHW still considers that gender dysphoria rather than gender identity should determine access to care and treatment. An urgent work thus remains, to clarify criteria under which adolescents with non-binary gender identity may be offered puberty-suppressing and gender-affirming hormonal treatment within a research framework.

References

1. Socialstyrelsen. God vård av barn och ungdomar med könsdysfori. Nationellt kunskapsstöd. https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/kunskapsstod/2015-4-6.pdf; 2015.
2. Statens beredning för medicinsk och social utvärdering. Hormonbehandling vid könsdysfori - barn och unga. En systematisk översikt och utvärdering av medicinska aspekter: SBU; 2022.
3. Littman L. Individuals Treated for Gender Dysphoria with Medical and/or Surgical Transition Who Subsequently Detransitioned: A Survey of 100 Detransitioners. Arch Sex Behav. 2021; 50(8):3353-69.
4. Socialstyrelsen. Utvecklingen av diagnosen könsdysfori - förekomst, samtidiga psykiatriska diagnoser och dödlighet i suicid. https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/ovrigt/2020-2-6600.pdf.; 2020.
5. Cohen-Kettenis PT, van Goozen SH. Sex reassignment of adolescent transsexuals: a follow-up study. J Am Acad Child Adolesc Psychiatry. 1997; 36(2):263-71.
6. Smith YL, van Goozen SH, Cohen-Kettenis PT. Adolescents with gender identity disorder who were accepted or rejected for sex reassignment surgery: a prospective follow-up study. J Am Acad Child Adolesc Psychiatry. 2001; 40(4):472-81.
7. Delemarre-van de Waal H, Cohen-Kettenis P. Clinical management of gender identity disorder in adolescents: a protocol on

psychological and paediatric endocrinology aspects. European Journal of Endocrinology. 2006; 155:S131-7.

8. Schagen SE, Cohen-Kettenis PT, Delemarre-van de Waal HA, Hannema SE. Efficacy and Safety of Gonadotropin-Releasing Hormone Agonist Treatment to Suppress Puberty in Gender Dysphoric Adolescents. J Sex Med. 2016; 13(7):1125-32.

OPEN ACCESS Freely available online PLoS one

Long-Term Follow-Up of Transsexual Persons Undergoing Sex Reassignment Surgery: Cohort Study in Sweden

Cecilia Dhejne[1], Paul Lichtenstein[2], Marcus Boman[2], Anna L. V. Johansson[2], Niklas Långström[2,3], Mikael Landén[1,2,4]*

1 Department of Clinical Neuroscience, Division of Psychiatry, Karolinska Institutet, Stockholm, Sweden, **2** Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, **3** Centre for Violence Prevention, Karolinska Institutet, Stockholm, Sweden, **4** Institute of Neuroscience and Physiology, The Sahlgrenska Academy at Gothenburg University, Gothenburg, Sweden

Abstract

Context: The treatment for transsexualism is sex reassignment, including hormonal treatment and surgery aimed at making the person's body as congruent with the opposite sex as possible. There is a dearth of long term, follow-up studies after sex reassignment.

Objective: To estimate mortality, morbidity, and criminal rate after surgical sex reassignment of transsexual persons.

Design: A population-based matched cohort study.

Setting: Sweden, 1973-2003.

Participants: All 324 sex-reassigned persons (191 male-to-females, 133 female-to-males) in Sweden, 1973–2003. Random population controls (10:1) were matched by birth year and birth sex or reassigned (final) sex, respectively.

Main Outcome Measures: Hazard ratios (HR) with 95% confidence intervals (CI) for mortality and psychiatric morbidity were obtained with Cox regression models, which were adjusted for immigrant status and psychiatric morbidity prior to sex reassignment (adjusted HR [aHR]).

Results: The overall mortality for sex-reassigned persons was higher during follow-up (aHR 2.8; 95% CI 1.8–4.3) than for controls of the same birth sex, particularly death from suicide (aHR 19.1; 95% CI 5.8–62.9). Sex-reassigned persons also had an increased risk for suicide attempts (aHR 4.9; 95% CI 2.9–8.5) and psychiatric inpatient care (aHR 2.8; 95% CI 2.0–3.9). Comparisons with controls matched on reassigned sex yielded similar results. Female-to-males, but not male-to-females, had a higher risk for criminal convictions than their respective birth sex controls.

Conclusions: Persons with transsexualism, after sex reassignment, have considerably higher risks for mortality, suicidal behaviour, and psychiatric morbidity than the general population. Our findings suggest that sex reassignment, although alleviating gender dysphoria, may not suffice as treatment for transsexualism, and should inspire improved psychiatric and somatic care after sex reassignment for this patient group.

Citation: Dhejne C, Lichtenstein P, Boman M, Johansson ALV, Långström N, et al. (2011) Long-Term Follow-Up of Transsexual Persons Undergoing Sex Reassignment Surgery: Cohort Study in Sweden. PLoS ONE 6(2): e16885. doi:10.1371/journal.pone.0016885

Editor: James Scott, The University of Queensland, Australia

Received September 30, 2010; **Accepted** January 9, 2011; **Published** February 22, 2011

Funding: Financial support was provided through the regional agreement on medical training and clinical research (ALF) between Stockholm County Council and the Karolinska Institutet, and through grants from the Swedish Medical Research Council (K2008-62x-14647-06-3) and the Royal Swedish Academy of Sciences (Torsten Amundson's Foundation). The sponsors of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. All authors had full access to the data in the study and the final responsibility for the decision to submit for publication was made by the corresponding author.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mikael.landen@neuro.gu.se

Introduction

Transsexualism (ICD-10),[1] or gender identity disorder (DSM-IV),[2] is a condition in which a person's gender identity - the sense of being a man or a woman - contradicts his or her bodily sex characteristics. The individual experiences gender dysphoria and desires to live and be accepted as a member of the opposite sex.

The treatment for transsexualism includes removal of body hair, vocal training, and cross-sex hormonal treatment aimed at making the person's body as congruent with the opposite sex as possible to alleviate the gender dysphoria. Sex reassignment also involves the surgical removal of body parts to make external sexual characteristics resemble those of the opposite sex, so called sex reassignment/confirmation surgery (SRS). This is a unique

intervention not only in psychiatry but in all of medicine. The present form of sex reassignment has been practised for more than half a century and is the internationally recognized treatment to ease gender dysphoria in transsexual persons.[3,4]

Despite the long history of this treatment, however, outcome data regarding mortality and psychiatric morbidity are scant. With respect to suicide and deaths from other causes after sex reassignment, an early Swedish study followed 24 transsexual persons for an average of six years and reported one suicide.[5] A subsequent Swedish study recorded three suicides after sex reassignment surgery of 175 patients.[6] A recent Swedish follow-up study reported no suicides in 60 transsexual patients, but one death due to complications after the sex reassignment surgery.[7] A Danish study reported death by suicide in 3 out of 29 operated male-to-female transsexual persons followed for an average of six years.[8] By contrast, a Belgian study of 107 transsexual persons followed for 4–6 years found no suicides or deaths from other causes.[9] A large Dutch single-centre study (N = 1,109), focusing on adverse events following hormonal treatment, compared the outcome after cross-sex hormone treatment with national Dutch standardized mortality and morbidity rates and found no increased mortality, with the exception of death from suicide and AIDS in male-to-females 25–39 years of age.[10] The same research group concluded in a recent report that treatment with cross-sex hormones seems acceptably safe, but with the reservation that solid clinical data are missing.[11] A limitation with respect to the Dutch cohort is that the proportion of patients treated with cross-sex hormones who also had surgical sex-reassignment is not accounted for.[10]

Data is inconsistent with respect to psychiatric morbidity post sex reassignment. Although many studies have reported psychiatric and psychological improvement after hormonal and/or surgical treatment,[7,12,13,14,15,16] other have reported on regrets,[17] psychiatric morbidity, and suicide attempts after SRS.[9,18] A recent systematic review and meta-analysis concluded that approximately 80% reported subjective improvement in terms of gender dysphoria, quality of life, and psychological symptoms, but also that there are studies reporting high psychiatric morbidity and suicide rates after sex reassignment.[19] The authors concluded though that the evidence base for sex reassignment "is of very low quality due to the serious methodological limitations of included studies."

The methodological shortcomings have many reasons. First, the nature of sex reassignment precludes double blind randomized controlled studies of the result. Second, transsexualism is rare [20] and many follow-ups are hampered by small numbers of subjects.[5,8,21,22,23,24,25,26,27,28] Third, many sex reassigned persons decline to participate in follow-up studies, or relocate after surgery, resulting in high drop-out rates and consequent selection bias.[6,9,12,21,24,28,29,30] Forth, several follow-up studies are hampered by limited follow-up periods.[7,9,21,22,26,30] Taken together, these limitations preclude solid and generalisable conclusions. A long-term population-based controlled study is one way to address these methodological shortcomings.

Here, we assessed mortality, psychiatric morbidity, and psychosocial integration expressed in criminal behaviour after sex reassignment in transsexual persons, in a total population cohort study with long-term follow-up information obtained from Swedish registers. The cohort was compared with randomly selected population controls matched for age and gender. We adjusted for premorbid differences regarding psychiatric morbidity and immigrant status. This study design sheds new light on transsexual persons' health after sex reassignment. It does not, however, address whether sex reassignment is an effective treatment or not.

Methods

National registers

The study population was identified by the linkage of several Swedish national registers, which contained a total of 13.8 million unique individuals. The Hospital Discharge Register (HDR, held by the National Board of Health and Welfare) contains discharge diagnoses, up to seven contributory diagnoses, external causes of morbidity or mortality, surgical procedure codes, and discharge date. Discharge diagnoses are coded according to the 8th (1969-1986), 9th (1987–1996), and 10th editions (1997-) of the International Classification of Diseases (ICD). The register covers virtually all psychiatric inpatient episodes in Sweden since 1973. Discharges that occurred up to 31 December 2003 were included. Surgical procedure codes could not be used for this study due to the lack of a specific code for sex reassignment surgery. The Total Population Register (TPR, held by Statistics Sweden) is comprised of data about the entire Swedish population. Through linkage with the Total Population Register it was possible to identify birth date and birth gender for all study subjects. The register is updated every year and gender information was available up to 2004/2005. The Medical Birth Register (MBR) was established in 1973 and contains birth data, including gender of the child at birth. National censuses based on mandatory self-report questionnaires completed by all adult citizens in 1960, 1970, 1980, and 1990 provided information on individuals, households, and dwellings, including gender, living area, and highest educational level. Complete migration data, including country of birth for immigrants for 1969–2003, were obtained from the TPR. In addition to educational information from the censuses, we also obtained highest educational level data for 1990 and 2000 from the Register of Education. The Cause of Death Register (CDR, Statistics Sweden) records all deaths in Sweden since 1952 and provided information on date of death and causes of death. Death events occurring up to 31 December 2003 are included in the study. The Crime Register (held by the National Council of Crime Prevention) provided information regarding crime type and date on all criminal convictions in Sweden during the period 1973–2004. Attempted and aggravated forms of all offences were also included. All crimes in Sweden are registered regardless of insanity at the time of perpetration; for example, for individuals who suffered from psychosis at the time of the offence. Moreover, conviction data include individuals who received custodial or non-custodial sentences and cases where the prosecutor decided to caution or fine without court proceedings. Finally, Sweden does not differ considerably from other members of the European Union regarding rates of violent crime and their resolution.[31]

Study population, identification of sex-reassigned persons (exposure assessment)

The study was designed as a population-based matched cohort study. We used the individual national registration number, assigned to all Swedish residents, including immigrants on arrival, as the primary key through all linkages. The registration number consists of 10 digits; the first six provide information of the birth date, whereas the ninth digit indicates the gender. In Sweden, a person presenting with gender dysphoria is referred to one of six specialised gender teams that evaluate and treat patients principally according to international consensus guidelines: Standards of Care.[3] With a medical certificate, the person applies to the National Board of Health and Welfare to receive permission for sex reassignment surgery and a change of legal sex status. A new national registration number signifying the new gender is assigned after sex reassignment surgery. The National

Board of Health and Welfare maintains a link between old and new national registration numbers, making it possible to follow individuals undergoing sex reassignment across registers and over time. Hence, sex reassignment surgery in Sweden requires (i) a transsexualism diagnosis and (ii) permission from the National Board of Health and Welfare.

A person was defined as exposed to sex reassignment surgery if two criteria were met: (i) at least one inpatient diagnosis of gender identity disorder diagnosis without concomitant psychiatric diagnoses in the Hospital Discharge Register, and (ii) at least one discrepancy between gender variables in the Medical Birth Register (from 1973 and onwards) or the National Censuses from 1960, 1970, 1980, or 1990 and the latest gender designation in the Total Population Register. The first criterion was employed to capture the hospitalization for sex reassignment surgery that serves to secure the diagnosis and provide a time point for sex reassignment surgery; the plastic surgeons namely record the reason for sex reassignment surgery, i.e., transsexualism, but not any co-occurring psychiatric morbidity. The second criterion was used to ensure that the person went through all steps in sex-reassignment and also changed sex legally.

The date of sex reassignment (start of follow-up) was defined as the first occurrence of a gender identity disorder diagnosis, without any other concomitant psychiatric disorder, in the Hospital Discharge Register after the patient changed sex status (any discordance in sex designation across the Censuses, Medical Birth, and Total Population registers). If this information was missing, we used instead the closest date in the Hospital Discharge Register on which the patient was diagnosed with gender identity disorder without concomitant psychiatric disorder prior to change in sex status. The reason for prioritizing the use of a gender identity disorder diagnosis *after* changed sex status over *before* was to avoid overestimating person-years at risk of sex-reassigned person.

Using these criteria, a total of 804 patients with gender identity disorder were identified, whereof 324 displayed a shift in the gender variable during the period 1973–2003. The 480 persons that did not shift gender variable comprise persons who either did not apply, or were not approved, for sex reassignment surgery. Moreover, the ICD 9 code 302 is a non specific code for sexual disorders. Hence, this group might also comprise persons that were hospitalized for sexual disorders other than transsexualism. Therefore, they were omitted from further analyses. Of the remaining 324 persons, 288 were identified with the gender identity diagnosis *after* and 36 *before* change of sex status. Out of the 288 persons identified *after* changed sex status, 185 could also be identified *before* change in sex status. The median time lag between the hospitalization *before* and *after* sex change for these 185 persons was 0.96 years (mean 2.2 years, SD 3.3).

Gender identity disorder was coded according to ICD-8: 302.3 (transsexualism) and 302.9 (sexual deviation NOS); ICD-9: 302 (overall code for sexual deviations and disorders, more specific codes were not available in ICD-9); and ICD-10: F64.0 (transsexualism), F64.1 (dual-role transvestism), F64.8 (other gender identity disorder), and F64.9 (gender identity disorder NOS). Other psychiatric disorders were coded as ICD-8: 290-301 and 303-315; ICD-9: 290-301 and 303-319; and ICD-10: F00-F63 as well as F65-F99.

Identification of population-based controls (unexposed group)

For each exposed person (N = 324), we randomly selected 10 unexposed controls. A person was defined as unexposed if there were no discrepancies in sex designation across the Censuses, Medical Birth, and Total Population registers *and* no gender identity disorder diagnosis according to the Hospital Discharge Register. Control persons were matched by sex and birth year and had to be alive and residing in Sweden at the estimated sex reassignment date of the case person. To study possible gender-specific effects on outcomes of interest, we used two different control groups: one with the same sex as the case individual at birth (birth sex matching) and the other with the sex that the case individual had been reassigned to (final sex matching).

Outcome measures

We studied mortality, psychiatric morbidity, accidents, and crime following sex reassignment. More specifically, we investigated: (1) all-cause mortality, (2) death by definite/uncertain suicide, (3) death by cardiovascular disease, and (4) death by tumour. Morbidity included (5) any psychiatric disorder (gender identity disorders excluded), (6) alcohol/drug misuse and dependence, (7) definite/uncertain suicide attempt, and (8) accidents. Finally, we addressed court convictions for (9) any criminal offence and (10) any violent offence. Each individual could contribute with several outcomes, but only one event per outcome. Causes of death (Cause of Death Registry from 1952 and onwards) were defined according to ICD as suicide (ICD-8 and ICD-9 codes E950-E959 and E980-E989, ICD-10 codes X60-X84 and Y10-Y34); cardiovascular disease (ICD-8 codes 390-458, ICD-9 codes 390-459, ICD-10 codes I00-I99); neoplasms (ICD-8 and ICD-9 codes 140-239, ICD-10 codes C00-D48), any psychiatric disorder (gender identity disorders excluded); (ICD-8 codes 290-301 and 303-315, ICD-9 codes 290-301 and 303-319, ICD-10 codes F00-F63 and F65-F99); alcohol/drug abuse and dependence (ICD-8 codes 303-304, ICD-9 codes 303-305 (tobacco use disorder excluded), ICD-10 codes F10-F16 and F18-F19 (x5 excluded); and accidents (ICD-8 and ICD-9 codes E800-E929, ICD-10 codes V01-X59).

Any criminal conviction during follow-up was counted; specifically, violent crime was defined as homicide and attempted homicide, aggravated assault and assault, robbery, threatening behaviour, harassment, arson, or any sexual offense.[32]

Covariates

Severe psychiatric morbidity was defined as inpatient care according to ICD-8 codes 291, 295-301, 303-304, and 307; ICD-9 codes 291-292, 295-298, 300-301, 303-305 (tobacco use disorder excluded), 307.1, 307.5, 308-309, and 311; ICD-10 codes F10-F16, F18-F25, F28-F45, F48, F50, and F60-F62. Immigrant status, defined as individuals born abroad, was obtained from the Total Population Register. All outcome/covariate variables were dichotomized (i.e., affected or unaffected) and without missing values.

Statistical analyses

Each individual contributed person-time from study entry (for exposed: date of sex reassignment; for unexposed: date of sex reassignment of matched case) until date of outcome event, death, emigration, or end of study period (31 December 2003), whichever came first. The association between exposure (sex reassignment) and outcome (mortality, morbidity, crime) was measured by hazard ratios (HR) with 95% CIs, taking follow-up time into account. HRs were estimated from Cox proportional hazard regression models, stratified on matched sets (1:10) to account for the matching by sex, age, and calendar time (birth year). We present crude HRs (though adjusted for sex and age through matching) and confounder-adjusted HRs [aHRs] for all outcomes. The two potential confounders, immigrant status (yes/no) and history of severe psychiatric morbidity (yes/no) prior to sex

reassignment, were chosen based on previous research[18,33] and different prevalence across cases and controls (Table 1).

Gender-separated analyses were performed and a Kaplan-Meier survival plot graphically illustrates the survival of the sex reassigned cohort and matched controls (all-cause mortality) over time. The significance level was set at 0.05 (all tests were two-sided). All outcome/covariate variables were without missing values, since they are generated from register data, which are either present (affected) or missing (unaffected). The data were analysed using SAS version 9.1 (SAS Institute Inc., Cary, NC, USA).

Ethics

The data linking of national registers required for this study was approved by the IRB at Karolinska Institutet, Stockholm. All data were analyzed anonymously; therefore, informed consent for each individual was neither necessary nor possible.

Results

We identified 324 transsexual persons (exposed cohort) who underwent sex reassignment surgery and were assigned a new legal sex between 1973 and 2003. These constituted the sex-reassigned (exposed) group. Fifty-nine percent (N = 191) of sex-reassigned persons were male-to-females and 41% (N = 133) female-to-males, yielding a sex ratio of 1.4:1 (Table 1).

The average follow-up time for all-cause mortality was 11.4 (median 9.1) years. The average follow-up time for the risk of being hospitalized for any psychiatric disorder was 10.4 (median 8.1).

Characteristics prior to sex reassignment

Table 1 displays demographic characteristics of sex-reassigned and control persons prior to study entry (sex reassignment). There were no substantial differences between female-to-males and male-to-females regarding measured baseline characteristics. Immigrant status was twice as common among transsexual individuals compared to controls, living in an urban area somewhat more common, and higher education about equally prevalent. Transsexual individuals had been hospitalized for psychiatric morbidity other than gender identity disorder prior to sex reassignment about four times more often than controls. To adjust for these baseline discrepancies, hazard ratios adjusted for immigrant status and psychiatric morbidity prior to baseline are presented for all outcomes [aHRs].

Mortality

Table 2 describes the risks for selected outcomes during follow-up among sex-reassigned persons, compared to same-age controls of the same birth sex. Sex-reassigned transsexual persons of both genders had approximately a three times higher risk of all-cause mortality than controls, also after adjustment for covariates. Table 2

Table 1. Baseline characteristics among sex-reassigned subjects in Sweden (N = 324) and population controls matched for birth year and sex.

Characteristic at baseline	Sex-reassigned subjects (N = 324)	Birth-sex matched controls (N = 3,240)	Final-sex matched controls (N = 3,240)
Gender			
Female at birth, male after sex change	133 (41%)	1,330 (41%)	1,330 (41%)
Male at birth, female after sex change	191 (59%)	1,910 (59%)	1,910 (59%)
Average age at study entry [years] (SD, min-max)			
Female at birth, male after sex change	33.3 (8.7, 20–62)	33.3 (8.7, 20–62)	33.3 (8.7, 20–62)
Male at birth, female after sex change	36.3 (10.1, 21–69)	36.3 (10.1, 21–69)	36.3 (10.1, 21–69)
Both genders	35.1 (9.7, 20–69)	35.1 (9.7, 20–69)	35.1 (9.7, 20–69)
Immigrant status			
Female at birth, male after sex change	28 (21%)	118 (9%)	100 (8%)
Male at birth, female after sex change	42 (22%)	176 (9%)	164 (9%)
Both genders	70 (22%)	294 (9%)	264 (8%)
Less than 10 years of schooling prior to entry vs. 10 years or more			
Females at birth, males after sex change	49 (44%); 62 (56%)	414 (37%); 714 (63%)	407 (36%); 713 (64%)
Males at birth, females after sex change	61 (41%); 89 (59%)	665 (40%); 1,011 (60%)	595 (35%); 1,091 (65%)
All individuals with data	110 (42%); 151 (58%)	1,079 (38%); 1,725 (62%)	1,002 (36%); 1,804 (64%)
Psychiatric morbidity* prior to study entry			
Female at birth, male after sex change	22 (17%)	47 (4%)	42 (3%)
Male at birth, female after sex change	36 (19%)	76 (4%)	72 (4%)
Both genders	58 (18%)	123 (4%)	114 (4%)
Rural [vs. urban] living area prior to entry			
Female at birth, male after sex change	13 (10%)	180 (14%)	195 (15%)
Male at birth, female after sex change	20 (10%)	319 (17%)	272 (14%)
Both genders	33 (10%)	499 (15%)	467 (14%)

Note:
*Hospitalizations for gender identity disorder were not included.
doi:10.1371/journal.pone.0016885.t001

Table 2. Risk of various outcomes among sex-reassigned subjects in Sweden (N = 324) compared to population controls matched for birth year and birth sex.

	Number of events cases/ controls 1973–2003	Outcome incidence rate per 1000 person-years 1973–2003 (95% CI)		Crude hazard ratio (95% CI) 1973–2003	Adjusted* hazard ratio (95% CI) 1973–2003	Adjusted* hazard ratio (95% CI) 1973–1988	Adjusted* hazard ratio (95% CI) 1989–2003
		Cases	Controls				
Any death	27/99	7.3 (5.0–10.6)	2.5 (2.0–3.0)	2.9 (1.9–4.5)	2.8 (1.8–4.3)	3.1 (1.9–5.0)	1.9 (0.7–5.0)
Death by suicide	10/5	2.7 (1.5–5.0)	0.1 (0.1–0.3)	19.1 (6.5–55.9)	19.1 (5.8–62.9)	N/A	N/A
Death by cardiovascular disease	9/42	2.4 (1.3–4.7)	1.1 (0.8–1.4)	2.6 (1.2–5.4)	2.5 (1.2–5.3)	N/A	N/A
Death by neoplasm	8/38	2.2 (1.1–4.3)	1.0 (0.7–1.3)	2.1 (1.0–4.6)	2.1 (1.0–4.6)	N/A	N/A
Any psychiatric hospitalisation‡	64/173	19.0 (14.8–24.2)	4.2 (3.6–4.9)	4.2 (3.1–5.6)	2.8 (2.0–3.9)	3.0 (1.9–4.6)	2.5 (1.4–4.2)
Substance misuse	22/78	5.9 (3.9–8.9)	1.8 (1.5–2.3)	3.0 (1.9–4.9)	1.7 (1.0–3.1)	N/A	N/A
Suicide attempt	29/44	7.9 (5.5–11.4)	1.0 (0.8–1.4)	7.6 (4.7–12.4)	4.9 (2.9–8.5)	7.9 (4.1–15.3)	2.0 (0.7–5.3)
Any accident	32/233	9.0 (6.3–12.7)	5.7 (5.0–6.5)	1.6 (1.1–2.3)	1.4 (1.0–2.1)	1.6 (1.0–2.5)	1.1 (0.5–2.2)
Any crime	60/350	18.5 (14.3–23.8)	9.0 (8.1–10.0)	1.9 (1.4–2.5)	1.3 (1.0–1.8)	1.6 (1.1–2.4)	0.9 (0.6–1.5)
Violent crime	14/61	3.6 (2.1–6.1)	1.4 (1.1–1.8)	2.7 (1.5–4.9)	1.5 (0.8–3.0)	N/A	N/A

Notes:
*Adjusted for psychiatric morbidity prior to baseline and immigrant status.
‡Hospitalisations for gender identity disorder were excluded.
N/A Not applicable due to sparse data.
doi:10.1371/journal.pone.0016885.t002

separately lists the outcomes depending on when sex reassignment was performed: during the period 1973-1988 or 1989–2003. Even though the overall mortality was increased across both time periods, it did not reach statistical significance for the period 1989–2003. The Kaplan-Meier curve (Figure 1) suggests that survival of transsexual persons started to diverge from that of matched controls after about 10 years of follow-up. The cause-specific mortality from suicide was much higher in sex-reassigned persons, compared to matched controls. Mortality due to cardiovascular disease was moderately increased among the sex-reassigned, whereas the numerically increased risk for malignancies was borderline statistically significant. The malignancies were lung cancer (N = 3), tongue cancer (N = 1), pharyngeal cancer (N = 1), pancreas cancer (N = 1), liver cancer (N = 1), and unknown origin (N = 1).

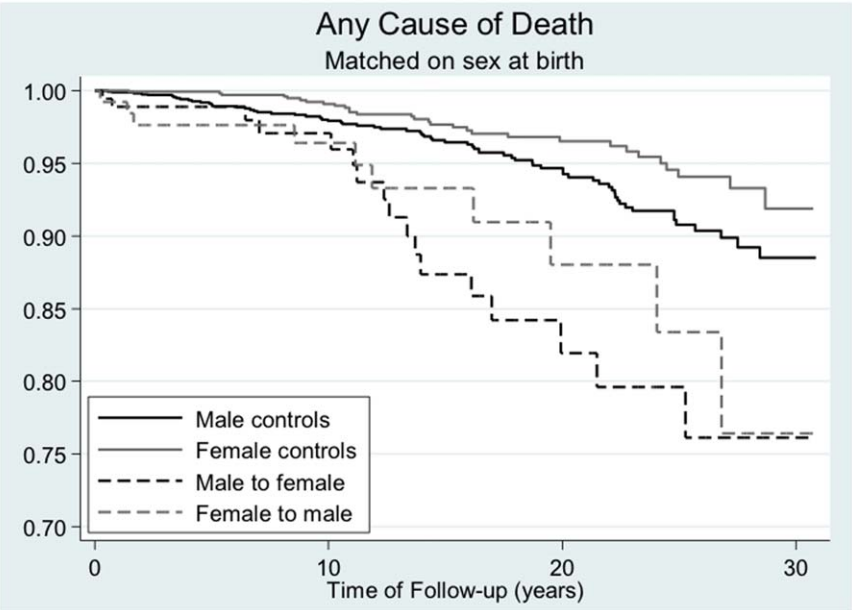


Figure 1. Death from any cause as a function of time after sex reassignment among 324 transsexual persons in Sweden (male-to-female: N = 191, female-to-male: N = 133), and population controls matched on birth year.
doi:10.1371/journal.pone.0016885.g001

Psychiatric morbidity, substance misuse, and accidents

Sex-reassigned persons had a higher risk of inpatient care for a psychiatric disorder other than gender identity disorder than controls matched on birth year and birth sex (Table 2). This held after adjustment for prior psychiatric morbidity, and was true regardless of whether sex reassignment occurred before or after 1989. In line with the increased mortality from suicide, sex-reassigned individuals were also at a higher risk for suicide attempts, though this was not statistically significant for the time period 1989–2003. The risks of being hospitalised for substance misuse or accidents were not significantly increased after adjusting for covariates (Table 2).

Crime rate

Transsexual individuals were at increased risk of being convicted for any crime or violent crime after sex reassignment (Table 2); this was, however, only significant in the group who underwent sex reassignment before 1989.

Gender differences

Comparisons of female-to-males and male-to-females, although hampered by low statistical power and associated wide confidence intervals, suggested mostly similar risks for adverse outcomes (Tables S1 and S2). However, violence against self (suicidal behaviour) and others ([violent] crime) constituted important exceptions. First, male-to-females had significantly increased risks for suicide attempts compared to both female (aHR 9.3; 95% CI 4.4–19.9) and male (aHR 10.4; 95% CI 4.9–22.1) controls. By contrast, female-to-males had significantly increased risk of suicide attempts only compared to male controls (aHR 6.8; 95% CI 2.1–21.6) but not compared to female controls (aHR 1.9; 95% CI 0.7–4.8). This suggests that male-to-females are at higher risk for suicide attempts after sex reassignment, whereas female-to-males maintain a female pattern of suicide attempts after sex reassignment (Tables S1 and S2).

Second, regarding any crime, male-to-females had a significantly increased risk for crime compared to female controls (aHR 6.6; 95% CI 4.1–10.8) but not compared to males (aHR 0.8; 95% CI 0.5–1.2). This indicates that they retained a male pattern regarding criminality. The same was true regarding violent crime. By contrast, female-to-males had higher crime rates than female controls (aHR 4.1; 95% CI 2.5–6.9) but did not differ from male controls. This indicates a shift to a male pattern regarding criminality and that sex reassignment is coupled to increased crime rate in female-to-males. The same was true regarding violent crime.

Discussion

Principal findings and comparison with previous research

We report on the first nationwide population-based, long-term follow-up of sex-reassigned transsexual persons. We compared our cohort with randomly selected population controls matched for age and gender. The most striking result was the high mortality rate in both male-to-females and female-to males, compared to the general population. This contrasts with previous reports (with one exception[8]) that did not find an increased mortality rate after sex reassignment, or only noted an increased risk in certain subgroups.[7,9,10,11] Previous clinical studies might have been biased since people who regard their sex reassignment as a failure are more likely to be lost to follow-up. Likewise, it is cumbersome to track deceased persons in clinical follow-up studies. Hence, population-based register studies like the present are needed to improve representativity.[19,34]

The poorer outcome in the present study might also be explained by longer follow-up period (median >10 years) compared to previous studies. In support of this notion, the survival curve (Figure 1) suggests increased mortality from ten years after sex reassignment and onwards. In accordance, the overall mortality rate was only significantly increased for the group operated before 1989. However, the latter might also be explained by improved health care for transsexual persons during 1990s, along with altered societal attitudes towards persons with different gender expressions.[35]

Mortality due to cardiovascular disease was significantly increased among sex reassigned individuals, albeit these results should be interpreted with caution due to the low number of events. This contrasts, however, a Dutch follow-up study that reported no increased risk for cardiovascular events.[10,11] A recent meta-analysis concluded, however, that data on cardiovascular outcome after cross-sex steroid use are sparse, inconclusive, and of very low quality.[34]

With respect to neoplasms, prolonged hormonal treatment might increase the risk for malignancies,[36] but no previous study has tested this possibility. Our data suggested that the cause-specific risk of death from neoplasms was increased about twice (borderline statistical significance). These malignancies (see Results), however, are unlikely to be related to cross-hormonal treatment.

There might be other explanations to increased cardiovascular death and malignancies. Smoking was in one study reported in almost 50% by the male-to females and almost 20% by female-to-males.[9] It is also possible that transsexual persons avoid the health care system due to a presumed risk of being discriminated.

Mortality from suicide was strikingly high among sex-reassigned persons, also after adjustment for prior psychiatric morbidity. In line with this, sex-reassigned persons were at increased risk for suicide attempts. Previous reports [6,8,10,11] suggest that transsexualism is a strong risk factor for suicide, also after sex reassignment, and our long-term findings support the need for continued psychiatric follow-up for persons at risk to prevent this.

Inpatient care for psychiatric disorders was significantly more common among sex-reassigned persons than among matched controls, both before and after sex reassignment. It is generally accepted that transsexuals have more psychiatric ill-health than the general population prior to the sex reassignment.[18,21,22,33] It should therefore come as no surprise that studies have found high rates of depression,[9] and low quality of life[16,25] also after sex reassignment. Notably, however, in this study the increased risk for psychiatric hospitalisation persisted even after adjusting for psychiatric hospitalisation prior to sex reassignment. This suggests that even though sex reassignment alleviates gender dysphoria, there is a need to identify and treat co-occurring psychiatric morbidity in transsexual persons not only before but also after sex reassignment.

Criminal activity, particularly violent crime, is much more common among men than women in the general population. A previous study of all applications for sex reassignment in Sweden up to 1992 found that 9.7% of male-to-female and 6.1% of female-to-male applicants had been prosecuted for a crime.[33] Crime after sex reassignment, however, has not previously been studied. In this study, male-to-female individuals had a higher risk for criminal convictions compared to female controls but not compared to male controls. This suggests that the sex reassignment procedure neither increased nor decreased the risk for criminal offending in male-to-females. By contrast, female-to-males were at a higher risk for criminal convictions compared to female controls and did not differ from male controls, which suggests increased crime proneness in female-to-males after sex reassignment.

Strengths and limitations of the study

Strengths of this study include nationwide representativity over more than 30 years, extensive follow-up time, and minimal loss to follow-up. Many previous studies suffer from low outcome ascertainment,[6,9,21,29] whereas this study has captured almost the entire population of sex-reassigned transsexual individuals in Sweden from 1973–2003. Moreover, previous outcome studies have mixed pre-operative and post-operative transsexual persons,[22,37] while we included only post-operative transsexual persons that also legally changed sex. Finally, whereas previous studies either lack a control group or use standardised mortality rates or standardised incidence rates as comparisons,[9,10,11] we selected random population controls matched by birth year, and either birth or final sex.

Given the nature of sex reassignment, a double blind randomized controlled study of the result after sex reassignment is not feasible. We therefore have to rely on other study designs. For the purpose of evaluating whether sex reassignment is an effective treatment for gender dysphoria, it is reasonable to compare reported gender dysphoria pre and post treatment. Such studies have been conducted either prospectively[7,12] or retrospectively,[5,6,9,22,25,26,29,38] and suggest that sex reassignment of transsexual persons improves quality of life and gender dysphoria. The limitation is of course that the treatment has not been assigned randomly and has not been carried out blindly.

For the purpose of evaluating the safety of sex reassignment in terms of morbidity and mortality, however, it is reasonable to compare sex reassigned persons with matched population controls. The caveat with this design is that transsexual persons before sex reassignment might differ from healthy controls (although this bias can be statistically corrected for by adjusting for baseline differences). It is therefore important to note that the current study is only informative with respect to transsexuals persons health after sex reassignment; no inferences can be drawn as to the effectiveness of sex reassignment as a treatment for transsexualism. In other words, the results should not be interpreted such as sex reassignment *per se* increases morbidity and mortality. Things might have been even worse without sex reassignment. As an analogy, similar studies have found increased somatic morbidity, suicide rate, and overall mortality for patients treated for bipolar disorder and schizophrenia.[39,40] This is important information, but it does not follow that mood stabilizing treatment or antipsychotic treatment is the culprit.

Other facets to consider are first that this study reflects the outcome of psychiatric and somatic treatment for transsexualism provided in Sweden during the 1970s and 1980s. Since then, treatment has evolved with improved sex reassignment surgery, refined hormonal treatment,[11,41] and more attention to psychosocial care that might have improved the outcome. Second, transsexualism is a rare condition and Sweden is a small country (9.2 million inhabitants in 2008). Hence, despite being based on a comparatively large national cohort and long-term follow-up, the statistical power was limited. Third, regarding psychiatric morbidity after sex reassignment, we assessed inpatient psychiatric care. Since most psychiatric care is provided in outpatient settings (for which no reliable data were available), underestimation of the *absolute* prevalences was inevitable. However, there is no reason to believe that this would change the *relative risks* for psychiatric morbidity unless sex-reassigned transsexual individuals were more likely than matched controls to be admitted to hospital for any given psychiatric condition.

Finally, to estimate start of follow-up, we prioritized using the date of a gender identity disorder diagnosis *after* changed sex status over *before* changed sex status, in order to avoid overestimating person-years at risk after sex-reassignment. This means that adverse outcomes might have been underestimated. However, given that the median time lag between the hospitalization before and after change of sex status was less than a year (see Methods), this maneuver is unlikely to have influenced the results significantly. Moreover, all deaths will be recorded regardless of this exercise and mortality hence correctly estimated.

Conclusion

This study found substantially higher rates of overall mortality, death from cardiovascular disease and suicide, suicide attempts, and psychiatric hospitalisations in sex-reassigned transsexual individuals compared to a healthy control population. This highlights that post surgical transsexuals are a risk group that need long-term psychiatric and somatic follow-up. Even though surgery and hormonal therapy alleviates gender dysphoria, it is apparently not sufficient to remedy the high rates of morbidity and mortality found among transsexual persons. Improved care for the transsexual group after the sex reassignment should therefore be considered.

Supporting Information

Table S1 Risk of various outcomes in sex-reassigned persons in Sweden compared to population controls matched for birth year and *birth sex*.
(DOCX)

Table S2 Risk of various outcomes in sex-reassigned persons in Sweden compared to controls matched for birth year and *final sex*.
(DOCX)

Author Contributions

Conceived and designed the experiments: CD PL AJ NL ML. Performed the experiments: MB AJ. Analyzed the data: CD PL MB AJ NL ML. Contributed reagents/materials/analysis tools: PL NL AJ. Wrote the paper: CD PL MB AJ NL ML.

References

1. World Health Organization (1993) The ICD-10 Classification of Mental and Behavioural Disorders. Diagnostic criteria for research. Geneva: WHO.
2. American Psychiatric Association, ed (1994) Diagnostic and Statistical Manual of Mental Disorders. Washington, DC: APA.
3. Meyer W, Bockting W, Cohen-Kettenis P, Coleman E, DiCeglie D, et al. (2002) The Harry Benjamin International Gender Dysphoria Association's Standards of Care for Gender Identity Disorders, Sixth Version. Journal of Psychology & Human Sexuality 13: 1–30.
4. Cohen-Kettenis PT, Gooren LJG (1999) Transsexualism: A review of etiology, diagnosis and treatment. J Psychosom Res 46: 315–333.
5. Wålinder J, Thuwe I (1975) A social-psychiatric follow-up study of 24 sex-reassigned transsexuals. Göteborg, Sweden: Scandinavian University Books.
6. Eldh J, Berg A, Gustafsson M (1997) Long-term follow up after sex reassignment surgery. Scand J Plast Reconstr Surg Hand Surg 31: 39–45.
7. Johansson A, Sundbom E, Höjerback T, Bodlund O (2010) A five-year follow-up study of Swedish adults with gender identity disorder. Arch Sex Behav 39: 1429–1437.
8. Sørensen T, Hertoft P (1982) Male and female transsexualism: the Danish experience with 37 patients. ArchSex Behav 11: 133–155.
9. De Cuypere G, T'Sjoen G, Beerten R, Selvaggi G, De Sutter P, et al. (2005) Sexual and physical health after sex reassignment surgery. Arch Sex Behav 34: 679–690.
10. van Kesteren PJ, Asscheman H, Megens JA, Gooren LJ (1997) Mortality and morbidity in transsexual subjects treated with cross-sex hormones. Clin Endocrinol Oxf 47: 337–342.

11. Gooren LJ, Giltay EJ, Bunck MC (2008) Long-term treatment of transsexuals with cross-sex hormones: extensive personal experience. J Clin Endocrinol Metab 93: 19–25.
12. Smith YL, van Goozen SH, Cohen-Kettenis PT (2001) Adolescents with gender identity disorder who were accepted or rejected for sex reassignment surgery: a prospective follow-up study. J Am Acad Child Adolesc Psychiatry 40: 472–481.
13. Smith YL, Van Goozen SH, Kuiper AJ, Cohen-Kettenis PT (2005) Sex reassignment: outcomes and predictors of treatment for adolescent and adult transsexuals. Psychol Med 35: 89–99.
14. Leavitt F, Berger JC, Hoeppner JA, Northrop G (1980) Presurgical adjustment in male transsexuals with and without hormonal treatment. J Nerv Ment Dis 168: 693–697.
15. Cohen Kettenis PT, van Goozen SH (1997) Sex reassignment of adolescent transsexuals: a follow-up study. J Am Acad Child Adolesc Psychiatry 36: 263–271.
16. Newfield E, Hart S, Dibble S, Kohler L (2006) Female-to-male transgender quality of life. Qual Life Res 15: 1447–1457.
17. Landén M, Wålinder J, Hambert G, Lundström B (1998) Factors predictive of regret in sex reassignment. Acta Psychiatrica Scandinavica 97: 284–289.
18. Hepp U, Kraemer B, Schnyder U, Miller N, Delsignore A (2005) Psychiatric comorbidity in gender identity disorder. J Psychosom Res 58: 259–261.
19. Murad MH, Elamin MB, Garcia MZ, Mullan RJ, Murad A, et al. (2010) Hormonal therapy and sex reassignment: a systematic review and meta-analysis of quality of life and psychosocial outcomes. Clin Endocrinol (Oxf) 72: 214–231.
20. Landén M, Wålinder J, Lundström B (1996) Incidence and sex ratio of transsexualism in Sweden. Acta Psychiatrica Scandinavica 93: 261–263.
21. Lobato MI, Koff WJ, Manenti C, da Fonseca Seger D, Salvador J, et al. (2006) Follow-up of sex reassignment surgery in transsexuals: a Brazilian cohort. Arch Sex Behav 35: 711–715.
22. Bodlund O, Kullgren G (1996) Transsexualism-General outcome and prognostic factors. A five year follow-up study of 19 transsexuals in the process of changing sex. Arch Sex Behav 25: 303–316.
23. Lindemalm G, Körlin D, Uddenberg N (1986) Long-term follow-up of "sex change" in 13 male-to-female transsexuals. Arch Sex Behav 15: 187–210.
24. Rauchfleisch U, Barth D, Battegay R (1998) [Results of long-term follow-up of transsexual patients]. Nervenarzt 69: 799–805.
25. Kuhn A, Bodmer C, Stadlmayr W, Kuhn P, Mueller MD, et al. (2009) Quality of life 15 years after sex reassignment surgery for transsexualism. Fertil Steril 92: 1685–1689 e1683.
26. Zimmermann A, Zimmer R, Kovacs L, Einodshofer S, Herschbach P, et al. (2006) [Transsexuals' life satisfaction after gender transformation operations]. Chirurg 77: 432–438.
27. Rehman J, Lazer S, Benet AE, Schaefer LC, Melman A (1999) The reported sex and surgery satisfactions of 28 postoperative male-to- female transsexual patients. Arch Sex Behav 28: 71–89.
28. Hepp U, Klaghofer R, Burkhard-Kubler R, Buddeberg C (2002) [Treatment follow-up of transsexual patients. A catamnestic study]. Nervenarzt 73: 283–288.
29. Lawrence AA (2003) Factors associated with satisfaction or regret following male-to-female sex reassignment surgery. Arch Sex Behav 32: 299–315.
30. Kaube H, Biemer E (1991) [Results of sex change operations in 30 transsexual patients: psychosocial and sexual adaptation–surgical complications]. Handchir Mikrochir Plast Chir 23: 276–278.
31. Dolmén L (2001) Brottsligheten i olika länder (Criminality in different countries). Stockholm: Brottsförebyggande rådet (the Swedish National Council for Crime Prevention).
32. Fazel S, Grann M (2006) The population impact of severe mental illness on violent crime. Am J Psychiatry 163: 1397–1403.
33. Landén M, Wålinder J, Lundström B (1998) Clinical characteristics of a total cohort of female and male applicants for sex reassignment: a descriptive study. Acta Psychiatrica Scandinavica 97: 189–194.
34. Elamin MB, Garcia MZ, Murad MH, Erwin PJ, Montori VM (2010) Effect of sex steroid use on cardiovascular risk in transsexual individuals: a systematic review and meta-analyses. Clin Endocrinol (Oxf) 72: 1–10.
35. Landén M, Innala S (2000) Attitudes toward transsexualism in a Swedish national survey. Archives of Sexual Behavior 29: 375–388.
36. Mueller A, Gooren L (2008) Hormone-related tumors in transsexuals receiving treatment with cross-sex hormones. Eur J Endocrinol 159: 197–202.
37. Vujovic S, Popovic S, Sbutega-Milosevic G, Djordjevic M, Gooren L (2009) Transsexualism in Serbia: a twenty-year follow-up study. J Sex Med 6: 1018–1023.
38. Rehman J, Lazer S, Benet AE, Schaefer LC, Melman A (1999) The reported sex and surgery satisfactions of 28 postoperative male-to-female transsexual patients. Arch Sex Behav 28: 71–89.
39. Ösby U, Brandt L, Correia N, Ekbom A, Sparén P (2001) Excess mortality in bipolar and unipolar disorder in Sweden. Arch Gen Psychiatry 58: 844–850.
40. Tidemalm D, Langstrom N, Lichtenstein P, Runeson B (2008) Risk of suicide after suicide attempt according to coexisting psychiatric disorder: Swedish cohort study with long term follow-up. Bmj 337: a2205.
41. Toorians AW, Thomassen MC, Zweegman S, Magdeleyns EJ, Tans G, et al. (2003) Venous thrombosis and changes of hemostatic variables during cross-sex hormone treatment in transsexual people. J Clin Endocrinol Metab 88: 5723–5729.

to create "win-win" relationships. By extension, critics of competition maintain that the NHS should do the same. These developments have been reinforced by concerns about the increase in management costs associated with the introduction of competition.

Estimates suggest that the NHS reforms may have resulted in up to £1bn extra being spent on administration, although changes in definitions make it difficult to be precise. This is because of the need to employ staff to negotiate and monitor contracts and to deal with the large volumes of paperwork involved in the contracting system. Ministers have responded to these concerns by streamlining the organisation of the NHS and introducing tight controls over management costs. They have also encouraged the use of long term contracts in order to reduce the transaction costs of the new arrangements.

Out of the ashes of competition has arisen a different policy agenda. This owes less to a belief in market forces than a desire to use the NHS reforms to achieve other objectives. The current agenda centres on policies to improve the health of the population, give greater priority to primary care, raise standards through the patient's charter, and ensure that medical decisions are evidence based. These policies hinge on effective planning and coordination in the NHS and all have been made more salient by the separation of purchaser and provider roles on which the reforms are based.

In particular, the existence of health authorities able to take an independent view of the population's health needs without being beholden to particular providers has changed the way in which decisions are made. To this extent the organisational changes introduced in 1991 have served to refocus attention on those whom the NHS exists to serve, even though the effects were neither anticipated nor intended when the reforms were designed. Like a potter moulding clay, only in the process of creation has the shape of the product become apparent. The effect of this policy shift has been to open up common ground between Labour and the Conservatives, notwithstanding the differences that remain.

Yet before the obituary of competition is written, the consequences of a return to planning need to be thought through. The NHS was reformed precisely because the old command and control system had failed to deliver acceptable improvements in efficiency and quality, and the limitations of planning must also be acknowledged. While competition as a reforming strategy may have had its day, there are nevertheless elements of this strategy which are worth preserving. Not least, the stimulus to improve performance which arises from the threat that contracts may be moved to an alternative provider should not be lost. The middle way between planning and competition is a path called contestability. This recognises that health care requires cooperation between purchasers and providers and the capacity to plan developments on a long term basis. At the same time, it is based on the premise that performance may stagnate unless there are sufficient incentives to bring about continuous improvements. Some of these incentives may be achieved through management action or professional pressure, and some may derive from political imperatives.

In addition, there is the stimulus to improve performance which exists when providers know that purchasers have alternative options. This continues to be part of the psychology of NHS decision making, even though ministers seem reluctant to use the language of markets. It is, however, a quite different approach than competitive tendering for clinical services, which would expose providers to the rigours of the market on a regular basis.

The essence of contestability is that planning and competition should be used together, with contracts moving only when other means of improving performance have failed. Put another way, in a contestable health service it is the possibility that contracts may move that creates an incentive within the system, rather than the actual movement of contracts. Of course for this to be a real incentive then contracts must shift from time to time, but this is only one element in the process and not necessarily the most important. As politicians prepare their plans for the future it is this path that needs to be explored.

CHRIS HAM
Director

Health Services Management Centre,
Birmingham B15 2RT

1 Smith R. William Waldegrave: thinking beyond the new NHS. *BMJ* 1990;**301**:711-4.
2 Bottomley V. *The new NHS: continuity and change*. London: Department of Health, 1995.

Evidence based medicine: what it is and what it isn't

It's about integrating individual clinical expertise and the best external evidence

Evidence based medicine, whose philosophical origins extend back to mid-19th century Paris and earlier, remains a hot topic for clinicians, public health practitioners, purchasers, planners, and the public. There are now frequent workshops in how to practice and teach it (one sponsored by the *BMJ* will be held in London on 24 April); undergraduate[1] and postgraduate[2] training programmes are incorporating it[3] (or pondering how to do so); British centres for evidence based practice have been established or planned in adult medicine, child health, surgery, pathology, pharmacotherapy, nursing, general practice, and dentistry; the Cochrane Collaboration and Britain's Centre for Review and Dissemination in York are providing systematic reviews of the effects of health care; new evidence based practice journals are being launched; and it has become a common topic in the lay media. But enthusiasm has been mixed with some negative reaction.[4-6] Criticism has ranged from evidence based medicine being old hat to it being a dangerous innovation, perpetrated by the arrogant to serve cost cutters and suppress clinical freedom. As evidence based medicine continues to evolve and adapt, now is a useful time to refine the discussion of what it is and what it is not.

Evidence based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research. By individual clinical expertise we mean the proficiency and judgment that individual clinicians acquire through clinical experience and clinical practice. Increased expertise is reflected in many ways, but especially in more effective and efficient diagnosis and in the more thoughtful identification and compassionate use of individual patients' predicaments, rights, and preferences in making clinical decisions about their care. By best available external clinical evidence we mean clinically relevant research, often from the

basic sciences of medicine, but especially from patient centred clinical research into the accuracy and precision of diagnostic tests (including the clinical examination), the power of prognostic markers, and the efficacy and safety of therapeutic, rehabilitative, and preventive regimens. External clinical evidence both invalidates previously accepted diagnostic tests and treatments and replaces them with new ones that are more powerful, more accurate, more efficacious, and safer.

Good doctors use both individual clinical expertise and the best available external evidence, and neither alone is enough. Without clinical expertise, practice risks becoming tyrannised by evidence, for even excellent external evidence may be inapplicable to or inappropriate for an individual patient. Without current best evidence, practice risks becoming rapidly out of date, to the detriment of patients.

This description of what evidence based medicine is helps clarify what evidence based medicine is not. Evidence based medicine is neither old hat nor impossible to practice. The argument that "everyone already is doing it" falls before evidence of striking variations in both the integration of patient values into our clinical behaviour[7] and in the rates with which clinicians provide interventions to their patients.[8] The difficulties that clinicians face in keeping abreast of all the medical advances reported in primary journals are obvious from a comparison of the time required for reading (for general medicine, enough to examine 19 articles per day, 365 days per year[9]) with the time available (well under an hour a week by British medical consultants, even on self reports[10]).

The argument that evidence based medicine can be conducted only from ivory towers and armchairs is refuted by audits from the front lines of clinical care where at least some inpatient clinical teams in general medicine,[11] psychiatry (J R Geddes *et al,* Royal College of Psychiatrists winter meeting, January 1996), and surgery (P McCulloch, personal communication) have provided evidence based care to the vast majority of their patients. Such studies show that busy clinicians who devote their scarce reading time to selective, efficient, patient driven searching, appraisal, and incorporation of the best available evidence can practice evidence based medicine.

Evidence based medicine is not "cookbook" medicine. Because it requires a bottom up approach that integrates the best external evidence with individual clinical expertise and patients' choice, it cannot result in slavish, cookbook approaches to individual patient care. External clinical evidence can inform, but can never replace, individual clinical expertise, and it is this expertise that decides whether the external evidence applies to the individual patient at all and, if so, how it should be integrated into a clinical decision. Similarly, any external guideline must be integrated with individual clinical expertise in deciding whether and how it matches the patient's clinical state, predicament, and preferences, and thus whether it should be applied. Clinicians who fear top down cookbooks will find the advocates of evidence based medicine joining them at the barricades.

Some fear that evidence based medicine will be hijacked by purchasers and managers to cut the costs of health care. This would not only be a misuse of evidence based medicine but suggests a fundamental misunderstanding of its financial consequences. Doctors practising evidence based medicine will identify and apply the most efficacious interventions to maximise the quality and quantity of life for individual patients; this may raise rather than lower the cost of their care.

Evidence based medicine is not restricted to randomised trials and meta-analyses. It involves tracking down the best external evidence with which to answer our clinical questions. To find out about the accuracy of a diagnostic test, we need to find proper cross sectional studies of patients clinically suspected of harbouring the relevant disorder, not a randomised trial. For a question about prognosis, we need proper follow up studies of patients assembled at a uniform, early point in the clinical course of their disease. And sometimes the evidence we need will come from the basic sciences such as genetics or immunology. It is when asking questions about therapy that we should try to avoid the non-experimental approaches, since these routinely lead to false positive conclusions about efficacy. Because the randomised trial, and especially the systematic review of several randomised trials, is so much more likely to inform us and so much less likely to mislead us, it has become the "gold standard" for judging whether a treatment does more good than harm. However, some questions about therapy do not require randomised trials (successful interventions for otherwise fatal conditions) or cannot wait for the trials to be conducted. And if no randomised trial has been carried out for our patient's predicament, we must follow the trail to the next best external evidence and work from there.

Despite its ancient origins, evidence based medicine remains a relatively young discipline whose positive impacts are just beginning to be validated,[12 13] and it will continue to evolve. This evolution will be enhanced as several undergraduate, postgraduate, and continuing medical education programmes adopt and adapt it to their learners' needs. These programmes, and their evaluation, will provide further information and understanding about what evidence based medicine is and is not.

DAVID L SACKETT
Professor

NHS Research and Development Centre for Evidence Based Medicine,
Oxford Radcliffe NHS Trust,
Oxford OX3 9DU

WILLIAM M C ROSENBERG
Clinical tutor in medicine

Nuffield Department of Clinical Medicine,
University of Oxford,
Oxford

J A MUIR GRAY
Director of research and development

Anglia and Oxford Regional Health Authority,
Milton Keynes

R BRIAN HAYNES
Professor of medicine and clinical epidemiology

McMaster University,
Hamilton, Ontario
Canada

W SCOTT RICHARDSON
Clinical associate professor of medicine

University of Rochester School of Medicine and Dentistry,
Rochester, New York,
USA

1 British Medical Association. *Report of the working party on medical education.* London: BMA, 1995.
2 Standing Committee on Postgraduate Medical and Dental Education. *Creating a better learning environment in hospitals. 1. Teaching hospital doctors and dentists to teach.* London: SCOPME, 1994.
3 General Medical Council. *Education committee report.* London: GMC, 1994.
4 Grahame-Smith D. Evidence based medicine: Socratic dissent. *BMJ* 1995;**310**:1126-7.
5 Evidence based medicine; in its place [editorial]. *Lancet* 1995;**346**:785.
6 Correspondence. Evidence based medicine. *Lancet* 1995;**346**:1171-2.
7 Weatherall DJ: The inhumanity of medicine. *BMJ* 1994;**309**;1671-2.
8 House of Commons Health Committee. *Priority setting in the NHS: purchasing. First report sessions 1994-95.* London: HMSO, 1995. (HC 134-1.)
9 Davidoff F, Haynes B, Sackett D, Smith R. Evidence based medicine: a new journal to help doctors identify the information they need. *BMJ* 1995;**310**:1085-6.
10 Sackett DL. Surveys of self-reported reading times of consultants in Oxford, Birmingham, Milton-Keynes, Bristol, Leicester, and Glasgow. In: Rosenberg WMC, Richardson WS, Haynes RB, Sackett DL. *Evidence-based medicine.* London: Churchill Livingstone (in press).
11 Ellis J, Mulligan I, Rowe J, Sackett DL. Inpatient general medicine is evidence based. *Lancet* 1995;**346**:407-10.
12 Bennett RJ, Sackett DL, Haynes RB, Neufeld VR. A controlled trial of teaching critical appraisal of the clinical literature to medical students. *JAMA* 1987;**257**:2451-4.
13 Shin JH, Flaynes RB, Johnston ME. Effect of problem-based, self-directed undergraduate education on life-long learning. *Can Med Assoc J* 1993;**148**:969-76.

For details of the international conference on evidence based medicine to be held in London on Wednesday 24 April 1996, contact the BMA/BMJ Conference Unit, telephone 0171 383 6605, fax 0171 383 6663.

Transgender Health
Volume 5, Number 4, 2020
Mary Ann Liebert, Inc.
DOI: 10.1089/trgh.2020.0006

ORIGINAL ARTICLE

Consensus Parameter: Research Methodologies to Evaluate Neurodevelopmental Effects of Pubertal Suppression in Transgender Youth

Diane Chen,[1–4,*,†] John F. Strang,[5–9,†] Victoria D. Kolbuck,[1] Stephen M. Rosenthal,[10] Kim Wallen,[11] Deborah P. Waber,[12,13] Laurence Steinberg,[14] Cheryl L. Sisk,[15] Judith Ross,[16,17] Tomas Paus,[18–20] Sven C. Mueller,[21,22] Margaret M. McCarthy,[23] Paul E. Micevych,[24] Carol L. Martin,[25] Baudewijntje P.C. Kreukels,[26] Lauren Kenworthy,[5–9] Megan M. Herting,[27,28] Agneta Herlitz,[29] Ira R.J. Hebold Haraldsen,[30] Ronald Dahl,[31] Eveline A. Crone,[32] Gordon J. Chelune,[33] Sarah M. Burke,[32] Sheri A. Berenbaum,[34,35] Adriene M. Beltz,[36] Julie Bakker,[37] Lise Eliot,[38] Eric Vilain,[39–41] Gregory L. Wallace,[42] Eric E. Nelson,[43,44] and Robert Garofalo[1,4]

[1]Potocsnak Family Division of Adolescent and Young Adult Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois, USA.
[2]Pritzker Department of Psychiatry and Behavioral Health, Ann & Robert H. Lurie Children's Hospital of Chicago, Chicago, Illinois, USA.
Departments of [3]Psychiatry & Behavioral Sciences and [4]Pediatrics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA.
[5]Division of Neuropsychology, Children's National Medical Center, Washington, District of Columbia, USA.
[6]Center for Neuroscience, Children's Research Institute, Children's National Medical Center, Washington, District of Columbia, USA.
Departments of [7]Pediatrics, [8]Neurology, and [9]Psychiatry, George Washington University School of Medicine, Washington, District of Columbia, USA.
[10]Division of Endocrinology, Benioff Children's Hospital, University of California San Francisco, San Francisco, California, USA.
[11]Department of Psychology, Yerkes National Primate Research Center, Emory University, Atlanta, Georgia, USA.
[12]Department of Psychiatry, Boston Children's Hospital, Boston, Massachusetts, USA.
[13]Department of Psychiatry, Harvard Medical School, Boston, Massachusetts, USA.
[14]Department of Psychology, Temple University, Philadelphia, Pennsylvania, USA.
[15]Department of Psychology, Michigan State University, East Lansing, Michigan, USA.
[16]Nemours duPont Hospital for Children, Wilmington, Delaware, USA.
[17]Department of Pediatrics, Thomas Jefferson University, Philadelphia, Pennsylvania, USA.
[18]Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Ontario, Canada.
Departments of [19]Psychology and [20]Psychiatry, University of Toronto, Toronto, Ontario, Canada.
[21]Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium.
[22]Department of Personality, Psychological Assessment and Treatment, University of Deusto, Bilbao, Spain.
[23]Program in Neuroscience, Department of Pharmacology, University of Maryland School of Medicine, Baltimore, Maryland, USA
[24]David Geffen School of Medicine at UCLA, Los Angeles, California, USA.
[25]School of Social and Family Dynamics, Arizona State University, Tempe, Arizona, USA.
[26]Amsterdam UMC, Location VUmc, Department of Medical Psychology and Center of Expertise on Gender Dysphoria, Amsterdam, The Netherlands.
Departments of [27]Preventive Medicine and [28]Pediatrics, University of Southern California, Los Angeles, California, USA.
[29]Section of Psychology, Department of Clinical Neuroscience, Karolinska Institutet, Stockholm, Sweden.
[30]Centre for Cognitive Health in Brain Disease, Oslo University Hospital, Oslo, Norway.
[31]School of Public Health, University of California, Berkeley, Berkeley, California, USA.
[32]Department of Developmental and Educational Psychology, Brain and Development Research Center, Leiden University, Leiden, The Netherlands.
[33]Department of Neurology, University of Utah School of Medicine, Salt Lake City, Utah, USA.
Departments of [34]Psychology and [35]Pediatrics, The Pennsylvania State University, University Park, Pennsylvania, USA.
[36]Department of Psychology, University of Michigan, Ann Arbor, Michigan, USA.
[37]GIGA Neurosciences, Liège University, Liège, Belgium.
[38]Department of Neuroscience, Rosalind Franklin University of Medicine & Science, Chicago, Illinois, USA.
[39]Center for Genetic Medicine Research, Children's National Medical Center, Washington, District of Columbia, USA.
[40]Department of Genomics and Precision Medicine, George Washington University, Washington, District of Columbia, USA.
[41]Epigenetics, Data, & Politics at Centre National de la Recherche Scientifique, Paris, France.
[42]Department of Speech, Language, and Hearing Science, George Washington University, Washington, District of Columbia, USA.
[43]Center for Biobehavioral Health, The Research Institute, Nationwide Children's Hospital, Columbus, Ohio, USA.
[44]Department of Pediatrics, The Ohio State University College of Medicine, Columbus, Ohio, USA.
[†]These two authors are co-first authors.

*Address correspondence to: Diane Chen, PhD, Potocsnak Family Division of Adolescent and Young Adult Medicine, Ann & Robert H. Lurie Children's Hospital of Chicago, 225 East Chicago Avenue, Box 161B, Chicago, IL 60611-2605, USA, E-mail: dichen@luriechildrens.org

Abstract

Purpose: Pubertal suppression is standard of care for early pubertal transgender youth to prevent the development of undesired and distressing secondary sex characteristics incongruent with gender identity. Preliminary evidence suggests pubertal suppression improves mental health functioning. Given the widespread changes in brain and cognition that occur during puberty, a critical question is whether this treatment impacts neurodevelopment.

Methods: A Delphi consensus procedure engaged 24 international experts in neurodevelopment, gender development, puberty/adolescence, neuroendocrinology, and statistics/psychometrics to identify priority research methodologies to address the empirical question: is pubertal suppression treatment associated with real-world neurocognitive sequelae? Recommended study approaches reaching 80% consensus were included in the consensus parameter.

Results: The Delphi procedure identified 160 initial expert recommendations, 44 of which ultimately achieved consensus. Consensus study design elements include the following: a minimum of three measurement time points, pubertal staging at baseline, statistical modeling of sex in analyses, use of analytic approaches that account for heterogeneity, and use of multiple comparison groups to minimize the limitations of any one group. Consensus study comparison groups include untreated transgender youth matched on pubertal stage, cisgender (i.e., gender congruent) youth matched on pubertal stage, and an independent sample from a large-scale youth development database. The consensus domains for assessment includes: mental health, executive function/cognitive control, and social awareness/functioning.

Conclusion: An international interdisciplinary team of experts achieved consensus around primary methods and domains for assessing neurodevelopmental effects (i.e., benefits and/or difficulties) of pubertal suppression treatment in transgender youth.

Keywords: expert consensus; Delphi; puberty blockers; GnRHa; transgender; adolescents

Introduction

Standards of care established by the World Professional Association for Transgender Health[1] and the Endocrine Society[2] recommend pubertal suppression for gender dysphoric transgender youth during early puberty (i.e., Tanner stages 2–3).[3,4] Pubertal suppression is achieved through administration of gonadotropin-releasing hormone agonists (GnRHa). When administered in early puberty, GnRHa suppress endogenous sex hormone production and prevent the development of undesired and irreversible secondary sex characteristics, thereby minimizing distress associated with pubertal development incongruent with gender identity.[5] For youth who later decide to initiate estrogen/testosterone (gender-affirming hormones [GAH]) treatment to induce development of the desired secondary sex characteristics, pubertal suppression may minimize the need for more invasive, surgical interventions (e.g., facial and chest surgery). For youth who decide not to pursue GAH treatment, discontinuing GnRHa will reactivate the hypothalamic-pituitary-gonadal axis and endogenous puberty will resume.[6]

Three longitudinal studies have examined psychosocial outcomes in GnRHa-treated transgender youth; two (conducted by the same research group) followed a single cohort over time, immediately before initiating GAH ($N=70$)[7] and later in early adulthood after surgery for gender affirmation ($N=55$).[8] The third study compared groups of GnRHa-treated ($n=35$) and untreated ($n=36$) youth longitudinally.[9] Findings across these studies include significant reductions in depressive symptoms and improvement in overall psychosocial functioning in GnRHa-treated transgender youth. A fourth cross-sectional study compared adolescents diagnosed with gender dysphoria (GD), who were treated with GnRHa and close to starting GAH treatment ($n=178$), adolescents newly referred for GD evaluation ($n=272$), and cisgender adolescents recruited from the general population ($n=651$) on self-reported internalizing/externalizing problems, self-harm/suicidality, and peer relationships.[10] Before medical treatment, clinic-referred adolescents reported more internalizing problems and self-harm/suicidality and poorer peer relationships compared to age-equivalent peers. GnRHa-treated transgender adolescents had fewer emotional and behavioral problems than clinic-referred, untreated adolescents and had comparable or better psychosocial functioning than same-age

cisgender peers. In addition to studies of youth, the 2015 U.S. Transgender Survey included questions about past gender-affirming medical treatment, including pubertal suppression. These questions were asked retrospectively and linked to reported current and lifetime mental health.[11] Individuals who received pubertal suppression treatment ($n = 89$), when compared to those who wanted pubertal suppression, but did not receive it ($n = 3405$), had lower odds of endorsing lifetime suicidal ideation on the survey. Given these five studies and the presumed reversibility of GnRHa treatment, pubertal suppression is increasingly offered to early pubertal transgender youth. It is important to note that there has been only one longitudinal report of adult outcomes,[8] and questions remain regarding the potential for both positive *and* disruptive effects of pubertal suppression on neurodevelopment.[12–14]

The pubertal and adolescent period is associated with profound neurodevelopment, including trajectories of increasing capacities for abstraction and logical thinking,[15] integrative thinking (e.g., consideration of multiple perspectives),[16,17] and social thinking and competence.[18,19] During this period, there is a developmental shift toward greater exploration and novelty seeking,[20,21] salience of peer perspectives and interactions,[22] and accelerated development of passions/interests and identities.[23] These developments lay the groundwork for adult functioning.[18,24] At the level of the brain, several primary neurodevelopmental processes unfold during adolescence, including myelin development[25] and changes in neural connectivity[26]; synaptic pruning[27] and gray matter maturation[28,29]; changes in functional connectivity[30]; and maturation of the prefrontal cortex[31] and the "social brain" network.[19] Adolescent neurodevelopmental processes underlie mental health risks, resilience, and outcomes.[32,33]

Considerable research has addressed the effects of puberty-related hormones on neurodevelopment, including hormone manipulation studies in nonhuman animals and observational studies in humans. Animal studies demonstrate pubertal hormones exert broad neuronal influence, including effects on neurogenesis, differentiation, apoptosis, dendritic branching, spine density, and regional gray and white matter volumes.[30,34] Androgen and estrogen receptors are found in high density within the hypothalamus and amygdala, and are also present in the hippocampus, midbrain, cerebellum, and cerebral cortex of the rodent and monkey.[35–37] This widespread receptor distribution in rodents may explain the diverse effects of pubertal hormones on both reproductive and nonreproductive behaviors, including anxiety, scent-marking, and food guarding.[34] In human studies, pubertal progression has been linked to developmental changes in reward,[38] social,[39] and emotional processing[40] as well as cognitive/emotional control.[41] However, consensus regarding pubertal impacts at the neural level—such as puberty-associated changes observed in magnetic resonance imaging (MRI) measures—has been more difficult to achieve.[42] Distinct puberty-related neurodevelopmental trajectories have been differentiated by sex.[43]

The combination of animal neurobehavioral research and human behavior studies supports the notion that puberty may be a *sensitive* period for brain organization:[44–46] that is, a limited phase when developing neural connections are uniquely shaped by hormonal and experiential factors, with potentially lifelong consequences for cognitive and emotional health. Studies have linked early life adversity to early puberty onset[47] and early puberty onset to poorer mental health.[48] There is also some evidence to suggest that delayed puberty onset predicts slightly poorer adult functional outcomes.[49] Taken as a whole, the existing knowledge about puberty and the brain raises the possibility that suppressing sex hormone production during this period could alter neurodevelopment in complex ways—not all of which may be beneficial.

Two small studies have assessed impacts of pubertal suppression on neural and cognitive functioning in peripubertal transgender youth. Staphorsius et al. compared brain and behavioral responses of GnRHa-treated (8 transgender girls [birth-assigned male] and 12 transgender boys [birth-assigned female]) and untreated transgender youth (10 of each sex) during an executive function task.[50] No group differences were found in task load-related brain activation; GnRHa-treated transgender girls demonstrated poorer performance compared with untreated transgender boys and cisgender controls. Schneider et al. evaluated a single pubertal transgender girl undergoing GnRHa with MRI scans of white matter and cognitive assessments at baseline (before GnRHa initiation) and at 22 and 28 months of pubertal suppression treatment.[51] During follow-up, white matter fractional anisotropy (i.e., a measure of axonal diameter, fiber coherence, and myelination) did not increase in the manner otherwise expected during puberty. By 22 months of pubertal suppression treatment, working memory scores dropped by more than half a standard deviation.

Larger-scale, longitudinal studies are required to understand possible neurodevelopmental impacts of pubertal suppression over time in transgender youth. Suppressing puberty may reduce dysphoria and diminish risks for poor mental health in this population, thereby exerting neuroprotective effects. If pubertal suppression disrupts aspects of neurodevelopment, it is possible these effects are temporary, with youth "catching up" developmentally after transitioning to GAH treatment or discontinuing GnRHa. However, pubertal suppression may prevent key aspects of development during a sensitive period of brain organization. Neurodevelopmental impacts might emerge over time, akin to the "late effects" cognitive findings associated with certain oncology treatments.[52] In sum, GnRHa treatment might produce a myriad of *varied* impacts, both positive and disruptive.

The goal of this study was to develop a framework in which these questions could be asked, and ultimately answered. We identify priority research methodologies that can be used to address the empirical question of how pubertal suppression in transgender youth may affect neurodevelopment and real-world functioning. Given the complexity of neural development during the pubertal period and the novelty of developmental research with transgender youth, this study employed a Delphi consensus method to leverage international expertise in neurodevelopment, gender development, puberty/adolescence, neuroendocrinology, and statistics/psychometrics. By engaging a community of experts in an iterative consensus-building procedure, this study aimed to advance thinking about efficacious designs by moving beyond individual research efforts and single-discipline approaches.

Methods

The Delphi procedure is a reliable iterative research method for establishing expert agreement,[53,54] and has been used extensively to address health-related questions, particularly in emerging fields of clinical care.[55–57] In the first round of a two-round Delphi procedure, a key question is presented to experts, who remain anonymous to one another throughout the Delphi process. Each expert provides responses/solutions to the question, which are then combined and organized by the study team. In the Delphi round two, experts rate each proposed statement/solution according to the level of agreement. Responses reaching the *a priori* consensus criterion are included as consensus statements. Given its anonymous iterative nature, the Delphi method avoids problems of typical expert work groups (e.g., adhering to the perspectives of more senior workgroup experts, inflexibly defending ideas) and allows for interaction among larger groups of experts from diverse locations and disciplines through asynchronous communication.[58–60]

We employed a two-round Delphi procedure to obtain expert consensus regarding the most efficacious research design elements to address the following research question: *What, if any, real-world impact does pubertal suppression have on transgender children's cognitive and neural development?* International experts in relevant research fields were identified and invited as follows:

1. An independent advisory panel consisting of five experts across key disciplines (see Acknowledgments section) was formed to identify international experts who, based on knowledge and experience, could best propose a research design to assess neurodevelopmental impacts of pubertal suppression in transgender youth.
2. Thirty-two recommended experts were vetted for their expertise; all met required criteria (i.e., a minimum of 10 first-author publications in relevant fields).
3. These experts were invited to participate in the Delphi procedure and were informed they would be invited to consider being a co-author of the resulting article. Twenty-eight experts responded: 20 agreed to participate, 4 declined due to lack of time, and 4 declined due to self-reported lack of expertise in this research area. Snowball sampling identified an additional 16 recommended experts, who were vetted (as described above) for their experience. Eight met criteria and were invited. Five of these experts participated, yielding a total of 25 experts agreeing to participate, 24 of whom completed the Delphi process. See Table 1 for academic institution locations and areas of expertise represented in the expert panel.

The Ann & Robert H. Lurie Children's Hospital of Chicago Institutional Review Board found that an expert Delphi consensus initiative did not require informed consent since the experts were direct partners in the research product. The first round of Delphi survey was distributed through the REDCap online survey platform and presented an overview of the research question with the following prompt for

Table 1. Institutional Representation and Self-Reported Areas of Expertise

	n
Location of academic institution	
United States	16
The Netherlands	3
Belgium	2
Canada	1
Norway	1
Sweden	1
Self-endorsed areas of expertise[a]	
Brain development	13
Adolescent development	12
Neuroendocrinology	11
Neuroimaging	11
Neuropsychology	8
Cognitive development	7
Developmental assessment	4
Expert in GnRHa	2
Other (write in)	4
Developmental social neuroscience	1
Transgender health	1
Genetics of sex chromosomes	1
Gender development	1

[a]Experts endorsed as many areas of expertise as applicable.
GnRHa, gonadotropin-releasing hormone agonists.

respondents: "What methods and tools should we use to identify clinically meaningful neurodevelopmental impacts of pubertal suppression? What type of longitudinal design and follow-ups are both practical and appropriate? What comparison groups might we consider?" This initial process yielded 131 distinct research design considerations; multiple descriptions of the same concept were collapsed into single statements. In the second Delphi round, each first-round research design consideration was presented back to the experts and rated as follows: a priority idea/approach or not a priority idea/approach. Experts could also select, "cannot rate due to lack of expertise." The first Delphi round also yielded lists of potential comparison groups and assessment domains (29 items). In the second Delphi round, participants were asked to rank order these items according to priority. For the priority rankings of comparison groups, the top-rated comparison group by each expert was given a value of 2 and the second rated comparison group was given a value of 1. A mean was calculated for each comparison group option based on these values and these mean scores were used to identify the overall priority rankings. For the list of priority domains to measure, a parallel approach was taken with the top 6 domains ranked by each expert.

All experts participated in the second Delphi round. Twenty-two of the Delphi experts participated in the construction of the resulting article and are co-authors listed in reverse alphabetical order by last name (authors 5–26). The Results section contains the exact statements endorsed as a "priority" approach by 80% or more of the Delphi panel.

Results

Four of the 131 individually presented statements were excluded from analyses because fewer than 15 experts rated them. Of the remaining 127 statements, 44 met the 80% or higher criterion for consensus and inclusion (see Table 2 for endorsement rates by statement). The average endorsement rate of included statements was 89.4%.

Consensus parameter

Study design considerations. A multicenter design with more than a single clinic will be necessary to recruit a sufficient sample size, as the effect size will likely be small. Meaningful effect sizes must be determined to ensure sufficient recruitment to power multiple expected comparisons accounting for attrition in a longitudinal design. Three time points of measurement are the absolute minimum. It will be necessary to manage the effects of repeated testing with a particular focus on minimizing the practice effects of a longitudinal design with multiple time points. For cognitive assessments, standardized batteries should be employed as: (1) there may be a larger database of norms available that the cohort could be compared to, in addition to a local comparison (control) group(s), (2) general composite scores within test batteries tend to provide more reliable and stable scores than individual tests, and (3) tasks within a category may be swapped in case of worries for learning effects. In any study of cognitive change based on serial assessments, reliability of measures is paramount (the consensus in the field is that tests should have a minimum test-retest reliability of >0.70). It may be pragmatic to use measures and methods from large representative studies, such as the Adolescent Brain Cognitive Development (ABCD) Study.

All processes being studied (e.g., gender identity, mental health, neural structure, and function) display considerable heterogeneity, and methods that fail to capture this will provide distorted findings and lead to biased clinical recommendations. Analyses based on group means (e.g., regression or ANOVAs) are unlikely to generalize to all individuals being treated. Therefore, it is necessary to collect enough data per person to characterize individual trajectories of change over time.

Table 2. Consensus Priority Recommendations Ordered by Consensus Ratings Within Categories

Study design considerations		
1	It would be helpful to follow these youth through and beyond initiation of cross-sex hormone treatment. Some aspects of human adolescent brain development are more related to pubertal hormone status than age *per se*, and to the extent that pubertal suppression may also put some features of brain development on hold; it would be good to know whether these features "catch up" once cross-sex hormone treatment has begun or whether a sensitive window for hormone-dependent brain development has closed.	22/22
2	Follow cohort after GnRHa treatment ends—collect data after the youth transition to GAH (when they complete their GnRHa treatment).	22/23
3	Any neurocognitive effect of GnRHa pubertal suppression may be complicated by the psychosocial and affective aspects of the transgender experience. This means that you would have to include multivariate models of both cognitive and psychosocial functioning.	22/23
4	Need to determine meaningful effect sizes and ensure sufficient statistical power for multiple expected comparisons with attrition.	21/22
5	Across the course of the study, three assessment points is the absolute minimum.	20/21
6	Need to use a multicenter design (not just one clinic).	21/23
7	Effects of GnRHa may not appear for several years. Any difference in brain structure due to GnRHa is likely to be seen over time (long term), rather than immediately.	20/22
8	Social and affective learning process may be affected by pausing puberty. These social and affective learning processes might cause subtle short-term differences that could ultimately cause clinically impactful and meaningful longer-term effects.	17/19
9	Of particular interest would be to also monitor the impact of hormonal therapy. One could then ask, "Does the trajectory change in response to cross-sex hormonal therapy or do they stay on the same trajectory as when they were on GnRHa?"	16/18
10	Assess target and comparison groups before puberty.	20/23
11	Need to manage the effects of repeated testing (i.e., minimize the practice effect of a longitudinal design with multiple time points).	19/22
12	The effect size will likely be small—therefore, you would need a large sample size.	19/23
13	The research design will need to account for the differences between youth who are assumed male versus assumed female as biological sex is differentially related to rate and pattern of cognitive development, connectome distinctiveness, and timing of peak brain volume.	19/23
14	All processes being studied (e.g., gender identity, mental health, and neural structure and function) display huge amounts of heterogeneity, and research methods that fail to capture this will provide distorted findings and lead to biased clinical recommendations. Analyses based on mean levels of these processes are unlikely to generalize to all individuals being treated (e.g., regressions or ANOVAs that compare groups with a slew of covariates). It is, therefore, necessary that enough data are collected per person to capture personalized trajectories of change across time. And the data need to be modeled in ways that reflect the heterogeneity of individual characteristics and trajectories.	18/22
Comparison groups and recruitment		
15	At least one control group should be cisgender participants as this area of research (i.e., hormones and the adolescent brain) is still rather new and more data are needed on all youth during this stage.	20/22
16	Critical to match the groups carefully to allow for evaluation of the effects of repeated testing (practice effects).	20/22
17	Comparison groups should be matched for pubertal stage.	19/21
18	Recruit all gender dysphoric youth across the pubertal age range, including those who are treated with GnRHa and those who are not.	18/21
19	This is not dissimilar from issues of discerning differences in cognitive trajectories in normal aging versus neurodegenerative disorders. The basic question involves cognitive growth curves among cisgender and transgender children overtime. There have been large-scale large-sample studies that have produced trajectories of brain development during the pre-pubertal, pubertal, and adolescent periods that could treated like a "brain growth curve."	15/18
20	Need more than one comparison group to minimize the limitations of any one comparison group (no single comparison group is ideal).	18/22
Pubertal staging/measurement		
21	Measure gonadal hormone levels.	23/23
22	Collect information on menstrual cycle and contraceptive use for female adolescents involved in the study.	23/23
23	Measure Tanner staging (i.e., secondary sex characteristics).	21/23
24	Measure height/weight.	18/22
Domains to measure		
25	Use white matter microstructure scans (diffusion tensor imaging)—and use a longitudinal imaging pipeline (which exists) for processing these data with scientific rigor.	15/15
26	A pragmatic methodological implication is to consider: (1) not only relying solely on measures of performance and behavior but also measures of learning and motivation, and (2) not only relying solely on measures of cognitive capacities but also on social, affective, and value-based learning processes.	19/20
27	If MRI is included, consider imaging approaches focused on the following domains: social-emotional processing, executive functioning, risk and reward processing, and self-concept.	20/22
28	Studies in laboratory rodents show that testosterone, acting during puberty, programs the ability to adapt behavior as a function of social experience—therefore, include instruments that evaluate social proficiency.	19/21
29	Use diffusion tensor imaging to analyze white matter at the microstructural level.	17/19

(*continued*)

Table 2. (Continued)

Study design considerations		
30	Studies in laboratory rodents show that ovarian hormones, acting during puberty, program cognitive flexibility by exerting long-lasting effects on excitatory-inhibitory balance in prefrontal cortex—so include instruments that evaluate behavioral flexibility.	18/21
31	Examine white matter development, which is important for processing speed.	17/20
32	Important to measure emotional functioning because it is bidirectionally related to executive functioning.	16/19
33	Look at white matter characteristics since they seem to develop during puberty under the influence of sex hormones.	15/18
34	One cannot study everything or study everything well. It will be critical to identify the priorities in such a study, as there is a danger of doing too much here. Consider the outcomes that matter most and the hypothesized mediating mechanisms. Focus on the outcomes of interest.	19/23
35	There is no clear evidence that progressing through puberty later than peers is associated with delayed maturation of abstract reasoning, executive function, and social capacities.	18/22
36	Use structural MRI (T1/T2)—and use a longitudinal imaging pipeline (which exists)—for processing these data with scientific rigor.	13/16
37	There is an emerging shift in thinking about the increase in reward sensitivity and sensation-seeking during puberty as related to social value learning. Dopamine release is not primarily a "reward" signal, but rather a learning signal (e.g., prediction error signal)—the natural increased salience of social learning (status, prestige, being admired, respected, liked, etc.) These pubertal changes may have small effects on immediate behavior, yet that could contribute to changes in patterns of behavior over time, which could lead to large individual differences in developmental trajectories for people, such as if they had blocked puberty.	13/16
Measurement approaches		
38	In any study of cognitive change based on serial assessments, reliability of the measure is paramount. The consensus in the field is that tests should have a minimum test-retest reliability of >0.70.	20/20
39	Behavioral measurements should include standardized measures appropriate for repeated assessment with high test-retest reliability.	21/22
40	Match acquisition parameters between imaging sites.	17/18
41	Consider implementing measures and methods from large representative protocols, such as the ABCD.	17/18
42	Neuroimaging should parallel the behavioral study—neural measures should be linked to neurocognitive and behavioral measures.	19/22
43	For cognitive assessment, use a standardized battery for two reasons: (1) there might be a larger database of norms available that the cohort could be compared to, in addition to the likely to be small comparison ("control") group, and (2) tasks within a category may be swapped in case of worries for learning effects.	18/21
44	Use "test batteries" that provide a general composite score as well as specific composites. By virtue of being composites, scores tend to be more reliable and stable than individual test scores.	17/20

The proportion represents the number of experts endorsing an item as a "priority" out of the total number of experts who rated the item as "priority" or "not priority." The denominator represents the number of experts rating an item as a "priority" or "not priority" (as opposed to "cannot rate due to lack of expertise" or skipping the item).

ABCD, Adolescent Brain Cognitive Development Study; GAH, gender-affirming hormones; MRI, magnetic resonance imaging.

Any GnRHa-induced neurocognitive effect may be complicated by psychosocial and affective aspects of the transgender experience. Therefore, multivariate models of both cognitive and psychosocial functioning should be included. Accounting for differences between birth-assigned male youth versus birth-assigned female youth is important, as sex is differentially related to the rate and pattern of cognitive development, connectome distinctiveness, and timing of peak brain volume. Assessments should begin before puberty in both treatment and comparison groups. The effects of pubertal suppression may not appear for several years. Any GnRHa-related difference in brain structure is likely to be observed over the long term, rather than immediately. Shifts in social and affective learning processes might cause subtle short-term differences that could ultimately result in clinically impactful longer-term effects. Therefore, studies should follow GnRHa-treated youth over time, including the time period after GnRHa treatment ends and/or when GAH commence. Some aspects of human adolescent brain development are more related to pubertal hormone status than age *per se*. To the extent that pubertal suppression may also put some features of brain development on hold, it is critical to know whether these features "catch up" (either once GAH treatment is initiated or if the adolescent elects to stop GnRHa and resume endogenous puberty), or whether a sensitive window for hormone-dependent brain development has closed. One way to measure this is to assess whether neurodevelopment shifts in response to initiating GAH following pubertal suppression: Do GnRHa-treated youth stay on the same neurodevelopmental trajectory as when puberty was suspended or does this trajectory change?

Comparison groups. To assess neurodevelopmental trajectories associated with GnRHa treatment, more than one comparison group is needed to minimize the limitations of any one comparison group. No single comparison group is ideal for this study question.

A rank order of possible comparison groups is provided in Table 3. Groups should also be well matched, given the effects of a repeated testing design (e.g., practice effects). Matching for pubertal/developmental stage will be critical, including Tanner staging, gonadal hormone levels, height and weight, and, among youth assigned female at birth, menstrual cycle and contraceptive use. A primary comparison should be between GnRHa-treated transgender youth and untreated transgender youth, but it will also be important to include comparisons with cisgender samples as research on hormones and the adolescent brain is still novel and emerging and more data are needed on all youth during this developmental period. One way to accomplish the latter is to employ existing large-scale databases from studies of brain development during the pre-pubertal, pubertal, and later-adolescent periods, treating them as brain growth curves for comparisons. This approach is similar to the differentiation of cognitive trajectories in normal aging versus neurodegenerative disorders. The basic research question involves comparing cognitive growth curves over time.

Domains to assess. It will be critical to prioritize assessment domains based on hypothesized mediating mechanisms, with the most important domains to measure as follows: mental/behavioral health, pubertal stage, executive function/control, gender identity/dysphoria, and social awareness/functioning. See Table 4 for a complete list of ranked domains. Although we (the Delphi experts) identify executive function/control and social functioning as key domains to measure, it is important to note that there is no clear evidence that progressing through puberty later than peers is associated with delayed maturation of abstract reasoning, executive function, and social capacities. Executive function and emotional functioning are bidirectionally related, and for this reason, the two should be integrated in models/analyses. In addition, cognitive/behavioral flexibility, a component of executive functioning, should be measured, given that studies in rodents show ovarian hormones, acting during puberty, program cognitive flexibility by exerting long-lasting effects on excitatory-inhibitory balance in the prefrontal cortex.[61] Studies in rodents also demonstrate that testosterone, acting during puberty, programs the ability to adapt behavior as a function of social experience.[34] Measurement approaches should extend beyond cognitive capacities alone, embedding social, affective, and value-based learning processes. There is an emerging shift in thinking about increases in reward sensitivity

Table 3. Rank Order of Priority Comparison Groups

Rank order of priority	Comparison group
1	Transgender youth who do not take GnRHa matched on pubertal status at the beginning of the study
2	Cisgender typically developing adolescents matched on pubertal status at the beginning of the study
3	Use a standardized battery and/or a large existing database of norms to compare to (in addition to a smaller comparison group)
4	Transgender youth who commence GnRHa treatment earlier compared to later in puberty
5	Siblings of transgender youth enrolling in the study (to serve as genetic and shared environmental controls)
6	Mixed clinical group of adolescents presenting for MH assessment/treatment in an outpatient setting matched on pubertal status
7[a]	Peers with mood disorders (to control for the overoccurrence of mental health distress in transgender youth) matched on pubertal status
7[a]	Youth with precocious puberty who are given GnRHa to delay puberty

This priority sequence was based on participants' top 2 ranked comparison groups, where the top rated comparison group was given a value of 2 and the second rated comparison group was given a value of 1. A mean score was derived for each comparison group based on participants' ratings and ordered from highest to lowest.

[a]Comparison groups received the same mean score in the ranking.

Table 4. Rank Order of Priority Domains of Characterization and Assessment

Rank order of priority	Domains of characterization and assessment
1	Mental/behavioral health (including suicidality/hopelessness)
2	Pubertal stage/development (Tanner staging/hormone levels)
3	Executive function/control and attention
4	Gender identity/dysphoria
5	Social awareness/functioning
6	Quality of life
7	Brain/functional connectivity
8	Brain structure/volume
9	Emotional awareness/functioning
10	Physical health symptoms and outcomes (especially in adulthood)
11	Adaptive/independence skills
12	General cognitive functioning (IQ)
13	Sensation seeking, risk taking, reward sensitivity, and motivation
14	Genetics (i.e., possible impacts of GnRHa on DNA and RNA expression)
15	Academic functioning
16	Processing speed
17	Memory systems

This priority sequence was based on participants' top 6 ranked domains to measure, where the top rated domain was given a value of 6 and the second rated comparison group was given a value of 5, and so on. A mean score was derived for each domain based on participants' ratings and ordered from highest to lowest.

and sensation-seeking during puberty as related to social-value learning.[18] Dopamine release is not primarily a "reward" signal, but rather a learning signal (e.g., prediction error signal)—the natural increased salience of social learning (e.g., status and prestige, being admired, respected, and liked). The effects of suspending puberty on the salience of social-value learning might produce small near-term effects, but could contribute to changes in patterns of behavior over time, leading to large individual differences in developmental trajectories for GnRHa-treated youth.

If neuroimaging is included, imaging approaches should focus on the following domains: social/emotional processing, executive functioning, risk and reward processing, and self-concept. Neuroimaging should parallel behavioral assessment. Neural measures should be linked to neurocognitive and behavioral measures. Acquisition parameters should be matched between imaging sites. Investigation of white matter development is important as myelination progresses during puberty, likely under the influence of sex hormones,[62] and is related to cognitive processing speed. Both structural MRI and diffusion tensor imaging approaches should be used for white matter imaging and analyzed using a longitudinal imaging pipeline for processing these data with scientific rigor.

Discussion

Puberty suppression has become an increasingly available option for transgender youth, and its benefits have been noted, particularly in the area of mental health. However, puberty is a major developmental process and the full consequences (both beneficial and adverse) of suppressing endogenous puberty are not yet understood. The experts who participated in this procedure believe the effects of pubertal suppression warrant further study, and this Delphi consensus process develops a framework from which future research endeavors can be built.

Expert consensus emphasized a minimum of three measurement time points, inclusion of multiple comparison groups to minimize the limitations of any one group, precision pubertal staging at baseline, accounting for sex in design and analysis, and the use of designs that capture heterogeneity in processes being studied. Focus on longer-term trajectories and outcomes was emphasized, given that effects of pubertal suppression on various processes may not be evident in the near term, and responses to delayed receipt of gonadal hormones may not be comparable to initial potentially organizing effects. Experts also highlighted that accounting for the psychosocial aspects of the transgender experience itself on development will require models that integrate both cognitive and psychosocial functioning. The highest endorsed measurement priorities were mental and behavioral health, executive function/cognitive control, and social awareness/functioning. The importance of interrelations between domains that mature during puberty/adolescence was also emphasized, including bidirectional relationships between cognitive and emotional control and links between reward sensitivity and social value learning. Regarding neuroimaging, experts stressed the importance of linking neural signatures to cognitive and behavioral measures, with attention to white matter development. Notably, while there was consensus in this approach to neuroimaging, there were divergent views as to whether a neuroimaging protocol should be prioritized in a study with limited resources. Some experts noted that insufficient work has been done on neural development during puberty in general and expending resources on an expensive neuroimaging protocol for this subset of youth may be premature, while others felt that defining underlying brain mechanisms by neuroimaging was important. Furthermore, at the final review of the article, four co-authors noted a concern with this specific Delphi consensus recommendation: "Accounting for differences between birth-assigned male youth versus birth-assigned female youth is important, as sex is differentially related to the rate and pattern of cognitive development, connectome distinctiveness, and timing of peak brain volume." The four authors felt that instead of "peak brain volume," a more appropriate measurement concept might be that of "structural brain metrics" (e.g., thickness and regional volumes).

Twelve different comparison groups were proposed in the first round of the Delphi and 8 of the 12 groups were rated as either first or second priority by at least 1 expert in the second Delphi round. This heterogeneity underscores the complexity of selecting comparison groups for this research and lends support to the experts' recommendation to engage more than one comparison group. The highest rated comparison groups were untreated transgender youth matched on pubertal stage, cisgender youth matched on pubertal stage, and a sample from a large-scale quasi-normative database (e.g., from the ABCD study) used as a "brain growth curve." These comparison groups are not without weaknesses. Untreated transgender youth may differ in their

intensity or experience of GD, level of parent support (e.g., are the parents against GnRHa treatment?), and socioeconomic status of the family and access to treatment (e.g., insurance coverage). A cisgender comparison group would lack gender-minority experience and associated stress.

Some statements approached, but did not reach consensus. For example, many experts suggested continuing assessments of transgender youth through young adulthood (mid-20s) when prefrontal development is near completion. Assessing adaptive functioning (everyday skills) over time due to the bidirectional link between executive functioning and adaptive behaviors was also often endorsed.

Not all relevant study considerations were raised by the Delphi panel. Neurodevelopmental impacts of pubertal suppression in transgender youth with neurodevelopmental differences/diagnoses (e.g., attention deficit/hyperactivity disorder and autism spectrum disorder) were not specifically addressed by the experts. Yet, evidence suggests an overoccurrence of neurodiversity characteristics (especially related to autism) among gender-referred youth.[55,63–66] The neurodevelopmental impacts of pubertal suppression on neurodiverse gender-diverse youth might well be different than in neurotypical gender-diverse youth, given variations in neurodevelopmental trajectories observed across neurodevelopmental conditions.[67–69]

This study included experts from a range of relevant disciplines—a strength and also a possible limitation. The varied disciplines allowed for a broader range of ideas and perspectives, but some specialized recommendations might not have been sufficiently understood by Delphi experts from other disciplines. It is possible that some useful recommendations were lost in the process because few experts had backgrounds relevant to them. In fact, four recommendations were dropped from consideration because more than nine experts indicated they could not rate the item or skipped the item. These four items included topics related to advanced growth curve modeling, impact of GnRHa on immune system functioning, multifactorial relationships between GD and neurodevelopment, and challenges associated with using alternative forms of measures in longitudinal designs. The Delphi team included experts across the fields of neuroscience, neurodevelopment, developmental measurement, and gender development; however, most were not specialized in clinical transgender care *per se*. This reflects the dearth of transgender care clinicians/specialists with research productivity in adolescent neurodevelopment. Thus, the experts could comment with authority on neurodevelopment, including gender development/dysphoria aspects of study design, but the real-world clinical care considerations may well be underdeveloped in the proposed research design. For example, the everyday lived experience of transgender youth seeking gender-affirming medical care would be unfamiliar to most neurodevelopmental researchers. After the Delphi procedure was completed, one panelist commented that pubertal hormones might play a role in organizing neurodevelopmental gender-related trajectories, including identity itself, which would be important to consider for a developmental study of gender diverse youth.

Despite these limitations, an international expert team successfully completed an iterative Delphi procedure achieving consensus around priority research design elements to study neurodevelopmental impacts of pubertal suppression in transgender youth. The resulting consensus parameter addresses broad design issues, including comparison groups, longitudinal design, neurodevelopmental targets for assessment, and measurement approaches. While it may not be possible to incorporate all consensus methodologies into a single study, this parameter may serve as a roadmap for a range of research initiatives investigating pubertal suppression treatment in transgender youth.

Acknowledgments
The advisory board was led by pediatric endocrinologist, Stephen Rosenthal, MD, and also included pediatric neuropsychologist, Gerard Gioia, PhD; gender-specialized neuroscientist, Lise Eliot, PhD; gender-specialized geneticist, Eric Vilain, MD, PhD; and developmental neuroscientist, Gregory E. Wallace, PhD. We would also like to recognize the contributions of Arthur P. Arnold, PhD, and Nicholas Allen, PhD, as experts on the Delphi panel.

Author Disclosure Statement
No competing financial interests exist.

Funding Information
This work was supported by a grant from the Ann & Robert H. Lurie Children's Hospital of Chicago Foundation funded by a grateful family that has chosen to remain anonymous. Study sponsors had no role in (1) study design, (2) collection, analysis, and interpretation of data, (3) writing of the report, or (4) the decision to submit the article for publication.

References

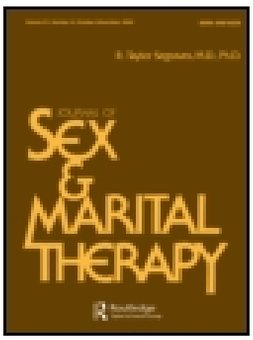
1. Coleman E, Bockting W, Botzer M, al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, Version 7. Int J Transgend. 2012;13:165–232.
2. Hembree WC, Cohen-Kettenis PT, Gooren L, et al. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2017;102:3869–3903.
3. Marshall WA, Tanner JM. Variations in the pattern of pubertal changes in boys. Arch Dis Child. 1970;45:13–23.
4. Marshall WA, Tanner JM. Variations in pattern of pubertal changes in girls. Arch Dis Child. 1969;44:291–303.
5. Kreukels BP, Cohen-Kettenis PT. Puberty suppression in gender identity disorder: the Amsterdam experience. Nat Rev Endocrinol. 2011;7:466–472.
6. de Vries AL, Cohen-Kettenis PT. Clinical management of gender dysphoria in children and adolescents: the Dutch approach. J Homosex. 2012;59:301–320.
7. de Vries AL, Steensma TD, Doreleijers TA, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. J Sex Med. 2011;8:2276–2283.
8. de Vries AL, McGuire JK, Steensma TD, et al. Young adult psychological outcome after puberty suppression and gender reassignment. Pediatrics. 2014;134:696–704.
9. Costa R, Dunsford M, Skagerberg E, et al. Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. J Sex Med. 2015;12:2206–2214.
10. van der Miesen AIR, Steensma TD, de Vries AL, et al. Psychological functioing in transgender adolescents before and after gender affirmative care compared to cisgender gender population peers. J Adolesc Health 2020. [Epub ahead of print]; DOI: 10.1016/j.jadohealth.2019.12.018.
11. Turban JL, King D, Carswell JM, Keuroghlian AS. Pubertal suppression for transgender youth and risk of suicidal ideation. Pediatrics. 2020;145:e20191725.
12. Costa R, Carmichael P, Colizzi M. To treat or not to treat: puberty suppression in childhood-onset gender dysphoria. Nat Rev Urol. 2016;13:456.
13. Vrouenraets LJJJ, Fredriks AM, Hannema SE, et al. Early medical treatment of children and adolescents with gender dysphoria: an empirical ethical study. J Adolesc Health. 2015;57:367–373.
14. Sadjadi S. The endocrinologist's office-puberty suppression: saving children from a natural disaster? J Med Humanit. 2013;34:255–260.
15. Dumontheil I. Development of abstract thinking during childhood and adolescence: the role of rostrolateral prefrontal cortex. Dev Cogn Neurosci. 2014;10:57–76.
16. Choudhury S, Blakemore SJ, Charman T. Social cognitive development during adolescence. Soc Cogn Affect Neurosci. 2006;1:165–174.
17. van den Bos W, van Dijk E, Westenberg M, et al. Changing brains, changing perspectives: the neurocognitive development of reciprocity. Psychol Sci. 2011;22:60–70.
18. Crone EA, Dahl RE. Understanding adolescence as a period of social-affective engagement and goal flexibility. Nat Rev Neurosci. 2012;13:636–650.
19. Kilford EJ, Garrett E, Blakemore SJ. The development of social cognition in adolescence: an integrated perspective. Neurosci Biobehav Rev. 2016;70:106–120.
20. Somerville LH, Sasse SF, Garrad MC, et al. Charting the expansion of strategic exploratory behavior during adolescence. J Exp Psychol Gen. 2017;146:155–164.
21. Steinberg L. Risk taking in adolescence: what changes, and why? Ann N Y Acad Sci. 2004;1021:51–58.
22. Albert D, Chein J, Steinberg L. Peer influences on adolescent decision making. Curr Dir Psychol Sci. 2013;22:114–120.
23. Dahl RE. Adolescent brain development: a period of vulnerabilities and opportunities. Keynote address. Ann N Y Acad Sci. 2004;1021:1–22.
24. McCormick EM, Telzer EH. Adaptive adolescent flexibility: neurodevelopment of decision-making and learning in a risky context. J Cogn Neurosci. 2017;29:413–423.
25. Paus T, Zijdenbos A, Worsley K, et al. Structural maturation of neural pathways in children and adolescents: in vivo study. Science. 1999;283:1908–1911.
26. Kolskar KK, Alnaes D, Kaufmann T, et al. Key brain network nodes show differential cognitive relevance and developmental trajectories during childhood and adolescence. eNeuro. 2018;5. DOI: 10.1523/ENEURO.0092-18.2018.
27. Selemon LD. A role for synaptic plasticity in the adolescent development of executive function. Transl Psychiatry. 2013;3:e238.
28. Gennatas ED, Avants BB, Wolf DH, et al. Age-related effects and sex differences in gray matter density, volume, mass, and cortical thickness from childhood to young adulthood. J Neurosci. 2017;37:5065–5073.
29. Wong AP, French L, Leonard G, et al. Inter-regional variations in gene expression and age-related cortical thinning in the adolescent brain. Cereb Cortex. 2018;28:1272–1281.
30. Goddings AL, Beltz A, Peper JS, et al. Understanding the role of puberty in structural and functional development of the adolescent brain. J Res Adolesc. 2019;29:32–53.
31. Crone EA, Steinbeis N. Neural perspectives on cognitive control development during childhood and adolescence. Trends Cogn Sci. 2017;21:205–215.
32. Paus T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? Nat Rev Neurosci. 2008;9:947–957.
33. Steinberg L. Cognitive and affective development in adolescence. Trends Cogn Sci. 2005;9:69–74.
34. Schulz KM, Sisk CL. Pubertal hormones, the adolescent brain, and the maturation of social behaviors: lessons from the Syrian hamster. Mol Cell Endocrinol. 2006;254:120–126.
35. Simerly RB, Chang C, Muramatsu M, Swanson LW. Distribution of androgen and estrogen receptor mRNA-containing cells in the rat brain: an in situ hybridization study. J Comp Neurol. 1990;294:76–95.
36. Shughrue PJ, Lane MV, Merchenthaler I. Comparative distribution of estrogen receptor-alpha and -beta mRNA in the rat central nervous system. J Comp Neurol. 1997;388:507–525.
37. Clark AS, MacLusky NJ, Goldman-Rakic PS. Androgen binding and metabolism in the cerebral cortex of the developing rhesus monkey. Endocrinology. 1988;123:932–940.
38. Braams BR, van Duijvenvoorde AC, Peper JS, Crone EA. Longitudinal changes in adolescent risk-taking: a comprehensive study of neural responses to rewards, pubertal development, and risk-taking behavior. J Neurosci. 2015;35:7226–7238.
39. Pfeifer JH, Kahn LE, Merchant JS, et al. Longitudinal change in the neural bases of adolescent social self-evaluations: effects of age and pubertal development. J Neurosci. 2013;33:7415–7419.
40. Spielberg JM, Olino TM, Forbes EE, Dahl RE. Exciting fear in adolescence: does pubertal development alter threat processing? Dev Cogn Neurosci. 2014;8:86–95.
41. Cservenka A, Stroup ML, Etkin A, Nagel BJ. The effects of age, sex, and hormones on emotional conflict-related brain response during adolescence. Brain Cogn. 2015;99:135–150.
42. Vijayakumar N, Op de Macks Z, Shirtcliff EA, Pfeifer JH. Puberty and the human brain: insights into adolescent development. Neurosci Biobehav Rev. 2018;92:417–436.
43. Ernst M, Benson B, Artiges E, et al. Pubertal maturation and sex effects on the default-mode network connectivity implicated in mood dysregulation. Transl Psychiatry. 2019;9:103.
44. Blakemore SJ, Mills KL. Is adolescence a sensitive period for sociocultural processing? Annu Rev Psychol. 2014;65:187–207.
45. Schulz KM, Sisk CL. The organizing actions of adolescent gonadal steroid hormones on brain and behavioral development. Neurosci Biobehav Rev. 2016;70:148–158.
46. Scherf KS, Smyth JM, Delgado MR. The amygdala: an agent of change in adolescent neural networks. Horm Behav. 2013;64:298–313.
47. Amir D, Jordan MR, Bribiescas RG. A longitudinal assessment of associations between adolescent environment, adversity perception, and economic status on fertility and age of menarche. PLoS One. 2016;11:e0155883.
48. Graber JA. Pubertal timing and the development of psychopathology in adolescence and beyond. Horm Behav. 2013;64:262–269.
49. Koerselman K, Pekkarinen T. Cognitive consequences of the timing of puberty. Labour Econ. 2018;54:1–13.
50. Staphorsius AS, Kreukels BP, Cohen-Kettenis PT, et al. Puberty suppression and executive functioning: an fMRI-study in adolescents with gender dysphoria. Psychoneuroendocrinology. 2015;56:190–199.

51. Schneider MA, Spritzer PM, Soll BMB, et al. Brain maturation, cognition and voice pattern in a gender dysphoria case under pubertal suppression. Front Hum Neurosci. 2017;11:528.
52. Rey-Casserly C, Diver T. Late effects of pediatric brain tumors. Curr Opin Pediatr. 2019;31:789–796.
53. Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. BMC Med Res Methodol. 2005;5:37.
54. Rowe G, Wright G. The Delphi technique as a forecasting tool: issues and analysis. Int J Forecast. 1999;15:353–375.
55. Strang JF, Meagher H, Kenworthy L, et al. Initial clinical guidelines for co-occurring autism spectrum disorder and gender dysphoria or incongruence in adolescents. J Clin Child Adolesc Psychol. 2018;47:105–115.
56. Taylor A, Tallon D, Kessler D, et al. An expert consensus on the most effective components of cognitive behavioural therapy for adults with depression: a modified Delphi study. Cogn Behav Ther. 2020;49:242–255.
57. Vogel C, Zwolinsky S, Griffiths C, et al. A Delphi study to build consensus on the definition and use of big data in obesity research. Int J Obes (Lond). 2019;43:2573–2586.
58. Hsu C-C, Sandford BA. The Delphi technique: making sense of consensus. Pract Assess Res Eval. 2007;12:1–8.
59. Keeney S, McKenna H, Hasson F. The Delphi Technique in Nursing and Health Research. West Sussex, UK: John Wiley & Sons, 2010.
60. Yousuf MI. Using experts' opinions through Delphi technique. Pract Assess Res Eval. 2007;12:1–8.
61. Piekarski DJ, Johnson CM, Boivin JR, et al. Does puberty mark a transition in sensitive periods for plasticity in the associative neocortex? Brain Res. 2017;1654(Pt B):123–144.
62. Ladouceur CD, Peper JS, Crone EA, Dahl RE. White matter development in adolescence: the influence of puberty and implications for affective disorders. Dev Cogn Neurosci. 2012;2:36–54.
63. Hisle-Gorman E, Landis CA, Susi A, et al. Gender dysphoria in children with autism spectrum disorder. LGBT Health. 2019;6:95–100.
64. Strauss P, Cook A, Winter S, et al. Trans Pathways: The Mental Health Experiences and Care Pathways of Trans Young People. Perth, Australia: Telethon Kids Institute, 2017.
65. de Vries ALC, Noens ILJ, Cohen-Kettenis PT, et al. Autism spectrum disorders in gender dysphoric children and adolescents. J Autism Dev Disord. 2010;40:930–936.
66. Akgül GY, Ayaz AB, Yildirim B, Fis NP. Autistic traits and executive functions in children and adolescents with gender dysphoria. J Sex Marital Ther. 2018;44:619–626.
67. Lin HY, Perry A, Cocchi L, et al. Development of frontoparietal connectivity predicts longitudinal symptom changes in young people with autism spectrum disorder. Transl Psychiatry. 2019;9:86.
68. Pugliese CE, Anthony LG, Strang JF, et al. Longitudinal examination of adaptive behavior in autism spectrum disorders: influence of executive function. J Autism Dev Disord. 2016;46:467–477.
69. Shaw P, Malek M, Watson B, et al. Development of cortical surface area and gyrification in attention-deficit/hyperactivity disorder. Biol Psychiatry. 2012;72:191–197.

Cite this article as: Chen D, Strang JF, Kolbuck VD, Rosenthal SM, Wallen K, Waber DP, Steinberg L, Sisk CL, Ross J, Paus T, Mueller SC, McCarthy MM, Micevych PE, Martin CL, Kreukels BPC, Kenworthy L, Herting MM, Herlitz A, Hebold Haraldsen IRJ, Dahl R, Crone EA, Chelune GJ, Burke SM, Berenbaum SA, Beltz AM, Bakker J, Eliot L, Vilain E, Wallace GL, Nelson EE, Garofalo R (2020) Consensus parameter: research methodologies to evaluate neurodevelopmental effects of pubertal suppression in transgender youth, *Transgender Health* 5:4, 246–257, DOI: 10.1089/trgh.2020.0006.

Abbreviations Used

ABCD = Adolescent Brain Cognitive Development
GAH = gender-affirming hormones
GD = gender dysphoria
GnRHa = gonadotropin-releasing hormone agonists
MRI = magnetic resonance imaging



Routledge
Taylor & Francis Group

Journal of Sex & Marital Therapy

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/usmt20

The Myth of "Reliable Research" in Pediatric Gender Medicine: A critical evaluation of the Dutch Studies—and research that has followed

E. Abbruzzese, Stephen B. Levine & Julia W. Mason

To cite this article: E. Abbruzzese, Stephen B. Levine & Julia W. Mason (2023): The Myth of "Reliable Research" in Pediatric Gender Medicine: A critical evaluation of the Dutch Studies—and research that has followed, Journal of Sex & Marital Therapy, DOI: 10.1080/0092623X.2022.2150346

To link to this article: https://doi.org/10.1080/0092623X.2022.2150346

Published online: 02 Jan 2023.

JOURNAL OF SEX & MARITAL THERAPY
https://doi.org/10.1080/0092623X.2022.2150346

Routledge
Taylor & Francis Group

ARTICLE COMMENTARY

The Myth of "Reliable Research" in Pediatric Gender Medicine: A critical evaluation of the Dutch Studies—and research that has followed

E. Abbruzzese[a], Stephen B. Levine[b] and Julia W. Mason[c]

[a]Society for Evidence-Based Gender Medicine, Twin Falls, ID, United States; [b]Department of Psychiatry, Case Western Reserve University, Cleveland, OH, United States; [c]Calcagno Pediatrics, Gresham, OH, United States

ABSTRACT

Two Dutch studies formed the foundation and the best available evidence for the practice of youth medical gender transition. We demonstrate that this work is methodologically flawed and should have never been used in medical settings as justification to scale this "innovative clinical practice." Three methodological biases undermine the research: (1) subject selection assured that only the most successful cases were included in the results; (2) the finding that "resolution of gender dysphoria" was due to the reversal of the questionnaire employed; (3) concomitant psychotherapy made it impossible to separate the effects of this intervention from those of hormones and surgery. We discuss the significant risk of harm that the Dutch research exposed, as well as the lack of applicability of the Dutch protocol to the currently escalating incidence of adolescent-onset, non-binary, psychiatrically challenged youth, who are preponderantly natal females. "Spin" problems—the tendency to present weak or negative results as certain and positive—continue to plague reports that originate from clinics that are actively administering hormonal and surgical interventions to youth. It is time for gender medicine to pay attention to the published objective systematic reviews and to the outcome uncertainties and definable potential harms to these vulnerable youth.

Introduction

In our recent paper on informed consent for youth gender transition, we recognized a serious problem: the field has a penchant for exaggerating what is known about the benefits of the practice, while downplaying the serious health risks and uncertainties (Levine et al., 2022a). As a result, a false narrative has taken root. It is that "gender-affirming" medical and surgical interventions for youth are as benign as aspirin, as well-studied as penicillin and statins, and as essential to survival as insulin for childhood diabetes—and that the vigorous scientific debate currently underway is merely "science denialism" motivated by ignorance, religious zeal, and transphobia (Drescher et al., 2022; McNamara et al., 2022; Turban, 2022). This highly politicized and fallacious narrative, crafted and promoted by clinician-advocates, has failed to withstand scientific scrutiny internationally, with public health authorities in Sweden, Finland, and most recently England doing a U-turn on pediatric gender transitions in the last 24 months (COHERE (Council for Choices in Health Care), 2020; Socialstyrelsen [National Board of Health and Welfare], 2022; National Health Service (NHS), 2022a). In the U.S., however, medical

CONTACT Stephen B. Levine sbl2@case.edu Department of Psychiatry, Case Western Reserve University, 23425 Commerce Park, STE 104, Cleveland, OH 44106-7078, United States

organizations so far have chosen to use their eminence to shield the practice of pediatric "gender affirmation" from scrutiny. In response to mounting legal challenges, these organizations have been exerting their considerable influence to insist the science is settled (American Medical Association (AMA), 2022). We argued that this stance stifles scientific debate, threatens the integrity and validity of the informed consent process—and ultimately, hurts the very patients it aims to protect.

To demonstrate problems in existing research, we discussed two seminal studies that gave rise to the now-common practice of performing gender transitions on young people by giving them puberty blockers, cross-sex hormones, and "gender-affirming" surgery (de Vries et al., 2011; de Vries et al., 2014). We argued that these Dutch studies suffer from such profound limitations that they should never have been used as justification for propelling these interventions into general medical practice. We called for rigorous clinical research into the interventions known as "gender-affirming" care before these interventions are further scaled. Until such research is available, we urged clinicians to disclose the profound uncertainties regarding the outcomes of this treatment pathway to enable patients and families to make better-informed decisions about their care.

Our assertions drew a response from the first author of these Dutch studies (de Vries, 2022).[1] de Vries dismissed much of our criticism as a mere "misunderstanding" of their gender clinic's process. While de Vries acknowledged some of the limitations in the Dutch research, she asserted that these gaps have since been sufficiently remedied by subsequent research from others in the field, rendering the practice of pediatric gender transition as proven beneficial, and ready to be widely scaled in general medical practice.

Having carefully examined de Vries' counterarguments, we failed to find a single instance where our "misunderstanding" could explain away the significant problems that we pointed out. In this article, we justify our position that neither the Dutch research, nor the research that followed, is fit for shaping policy or treatment decisions regarding gender dysphoric youth at the population level. We present our response to de Vries in three sections. *First,* we provide a more complete justification for our assertions of the significant flaws in the foundational Dutch research. *Second,* we demonstrate that the claims that subsequent research remedied the deficiencies in the prior research are untrue. *Third*, we provide recommendations for research structure to yield reliable, trustworthy information. We conclude with the sense of urgency to avoid future harms by reminding readers of the intrinsic value of high-quality science.

Before we embark on outlining the critical methodological limitations of the Dutch research, we would like to make it clear that it is not our intention to discredit the Dutch clinicians' past work. The quality of the Dutch studies, while unacceptably low by today's standards, is commensurate with clinical and research practices in the 1990s. The key problem in pediatric gender medicine is not the lack of research rigor in the *past*—it is the field's *present-day* denial of the profound problems in the existing research, and an unwillingness to engage in high quality research requisite in evidence-based medicine.

Evidence-based medicine vs empirical-based medicine

When the Dutch clinicians launched the practice of pediatric gender transition, it was not uncommon for medical professionals to practice medicine based on "empirical evidence," relying on expert opinion and often backed by only minimal research (Drisko & Friedman, 2019). The term "evidence-based medicine" and its focus on quality comparative clinical research to determine optimal treatment only emerged in the 1990s (Guyatt, 1993). The Dutch researchers began to medically transition gender dysphoric adolescents in the late 1980s–early 1990s—just as medicine was starting to undergo this major paradigm shift.

Examining the Dutch research from today's vantage point, their gender-transitioning of youth is most consistent with the "innovative practice" framework. This framework allows clinicians

to implement untested but promising interventions for a condition which, if left untreated, might have dire outcomes; when existing treatment options seem ineffective; and when the number of affected patients is small (Brierley & Larcher, 2009; Earl, 2019). The number of adolescents suffering from gender dysphoria in the 1990s was exceedingly small. Evidence was starting to demonstrate that gender reassignment undertaken in adulthood failed to resolve trans people's mental health problems (Cohen-Kettenis & Van Goozen, 1997). The Dutch clinicians hoped that the "less positive results among adults" (p. 266) would be remedied with early adolescent gender transition. In this context, the methodological deficiencies in the foundational Dutch research ought not to be viewed as a *failure*. It was never their goal to generate *reliable reproducible research*. In fact, the many irregularities, which we elucidate below, reflect the Dutch *success* at rapidly evolving their approaches to reach a point of *technical excellence*: convincing physical transformations of adolescent bodies that satisfied the young patients (Biggs, 2022). These clinicians were "flying the plane while building the plane," and their published research merely reflects this messy clinical reality.

The "innovative practice" model of care is a double-edged sword. On the one hand, it rapidly advances the medical field. On the other hand, it is capable of hurting individuals and societies by promoting a nonbeneficial or harmful intervention. For these reasons, it is an ethical requirement that as soon as viability of a new intervention is demonstrated under the "innovative practice" framework, the research must move into high-quality clinical research settings capable of demonstrating that the benefits outweigh the risks. This step is imperative because it prevents "runaway diffusion"—the phenomenon whereby the medical community mistakes a small innovative experiment as a proven practice, and a potentially nonbeneficial or harmful practice "escapes the lab," rapidly spreading into general clinical settings (Earl, 2019).

"Runaway diffusion" is exactly what has happened in pediatric gender medicine. "Affirmative treatment" with hormones and surgery rapidly entered general clinical practice worldwide, without the necessary rigorous clinical research to confirm the hypothesized robust and lasting psychological benefits of the practice. Nor was it ever demonstrated that the benefits were substantial enough to outweigh the burden of lifelong dependence on medical interventions, infertility and sterility, and various physical health risks. The studies also failed to quantify the risk to "false positives"—that is, those gender dysphoric youth whose distress would have remitted with time without resorting to irreversible medical and surgical interventions.

The speed of the "runaway diffusion" accelerated exponentially when pediatric gender dysphoria/transgender identity went from a relatively rare phenomenon before 2015, to one that impacts as many as 1 in 10–20 young people in the Western world (American College Health Association [ACHA], 2022; Johns et al., 2019; Kidd et al. 2021). The current politicization of transgender healthcare has provided further fuel to the rapid proliferation of youth gender reassignment. A proposal by the U.S. government to mandate healthcare entities to provide "gender-affirming" interventions to minors, or risk claims of "discrimination" and loss of federal healthcare funding is yet another example of "runaway diffusion" (Health and Human Services [HHS], 2022; Keith, 2022).

The difficult task of reversing runaway diffusion begins with a systematic review of evidence, follows with updating treatment guidelines, and culminates with de-implementation of unproven or harmful practices, known as "practice reversals" (Herrera-Perez et al., 2019; Prasad, 2011; Prasad & Ioannidis, 2014). *Systematic reviews of evidence* play a uniquely important role in this process. Rather than arbitrarily selecting studies and simply restating their results and conclusions, systematic reviews of evidence analyze *all of the available evidence* meeting pre-specified criteria and *scrutinize the studies* for methodological bias and errors, issuing an overarching conclusion about what's known about the effects of an intervention based on the totality of the evidence (Higgins et al., 2022). A "practice reversal" of pediatric gender transitions has already begun. Several recent international systematic reviews of evidence have concluded that the practice of pediatric gender transition rests on *low to very low quality evidence*—meaning that the benefits reported by the existing studies are unlikely to be true due to profound problems in the study designs (National

Institute for Health and Care Excellence (NICE), 2020a, 2020b; Pasternack et al., 2019; SBU (Swedish Agency for Health Technology Assessment and Assessment of Social Services), 2022). Following these systematic reviews of evidence, three European countries—Sweden, Finland and England—have begun to articulate new and much more cautious treatment guidelines for gender dysphoric youth, which prioritize noninvasive psychosocial interventions while sharply restricting the provision of hormones and surgery (COHERE (Council for Choices in Health Care), 2020; Socialstyrelsen [National Board of Health and Welfare], 2022; NHS, 2022a).

Paradoxically, this international reckoning has had almost no influence on the U.S. gender medicine establishment. When Florida's Medical Board, following an overview of existing systematic reviews (Brignardello-Peterson & Wiercioch, 2022), took on the question of regulating pediatric gender medicine and invited the proponents of pediatric gender transitions to reconcile their stance with the recent European developments, these clinician advocates were either unaware of the European changes, or minimized their extent and significance (Janssen, 2022 00:46:43; McNamara, 2022 01:45:27). More generally, when faced with questions about the rapidly growing numbers of youth subjected to highly invasive and often irreversible interventions based on *low to very low quality evidence*, the field of U.S. pediatric gender medicine has chosen to throw its weight behind two indefensible and contradictory claims: (1) that "low quality evidence" is a misleading technical term which actually describes high quality reliable research; and (2) that true high quality research can only come from randomized controlled trials, which are unattainable and unethical (Drescher, 2022; McNamara et al., 2022). We refuted these misleading claims in our recent publication (Levine et al., 2022b).

As we begin our discussion of the profound limitations in the two foundational Dutch studies that have propelled the practice of pediatric gender transition into mainstream clinical practice worldwide, we are aware that we are mounting a serious challenge to the research that has been viewed by many as the "gold standard" in the field. Questioning this assumption, we welcome further debate. A quote from philosopher Karl Popper, perceptively invoked by Balon (2022), is particularly apt: "the growth of knowledge depends entirely on disagreement."

I. The "Dutch studies" are deeply flawed

There is no argument that the Dutch experience, and in particular two Dutch studies—de Vries et al. (2011), and de Vries et al. (2014)—forms the foundation of the practice of youth gender transition. It is evident when examining prevailing treatment guidelines. The Endocrine Society's statements regarding the potential benefits of puberty blockers and cross-sex hormones in gender dysphoric adolescents are supported only by references to these two studies (Hembree et al., 2017, p. 12, p. 16). Similarly, the World Professional Association for Transgender Health (WPATH) "Standards of Care" guidelines version 7 (SOC 7)—the version under which the practice of medicalization of gender dysphoric youth became widespread—only references the Dutch experience (Coleman et al., 2012). Despite several newer studies available, the proponents of gender affirmation still correctly emphasize that "the best longitudinal data we have on transgender youth comes primarily out of the Dutch clinic…the Dutch studies in the Dutch model of care. That's the prevailing model that most of the American clinics have based their care upon" (Janssen, 2022, 00:47:42). de Vries in her response to us, also agrees with this: "…indeed, as of today, the Dutch papers, and especially the de Vries et al., 2014 study, are still used as main evidence for provision of early medical intervention including puberty blockers in transgender youth (de Vries et al., 2014)" (de Vries, 2022, p. 2).

The two main Dutch studies in question, de Vries et al., 2011, and de Vries et al., 2014 (from here on, "the Dutch studies") convincingly demonstrated that hormonal and surgical interventions can successfully change the phenotypical appearance of secondary sex characteristics of adolescents and young adults. What the studies *failed* to show, however, is that these physical changes resulted in meaningful psychological improvements significant enough to justify the adverse effects of the treatment—including the *certainty* of sterility.

Besides the lack of control group and a small final sample of 55 cases, with key outcomes available for as few as 32 individuals, there are *three major areas of concern* that render these studies unfit for clinical or policy decision-making.

A. High risk of bias: The Dutch studies suffer from multiple sources of bias which undermine confidence into the reported "benefits." The subject selection assured that only the most successful cases at each treatment stage were included in reported results. The linchpin finding of "resolution of gender dysphoria" is entirely invalid, since the home-grown gender dysphoria scale and its scoring mechanism were reversed after treatment, essentially guaranteeing a significant post-surgical drop in "gender dysphoria" scores. The finding of modest psychological benefits was compromised by the conflation of medical interventions with psychotherapy, making it impossible to determine whether gender reassignment, therapy, or the psychological maturation that occurs with the passage of time led to these few modest "improvements."
B. Incompleteness of evidence regarding physical health risks: The Dutch studies did not evaluate *physical health* outcomes of "gender-affirmative" treatments, even though adverse effects of hormonal interventions on bone and brain had been hypothesized from the start (and were confirmed by subsequent research). Even without setting out to assess the risks, the Dutch research inadvertently revealed that the rate of short-term morbidity and mortality associated with "gender-affirming" interventions may be as high as 6%-7%.
C. Poor generalizability/applicability to current cases: Today, most youth suffer from post-pubertal onset of gender dysphoria and significant mental illness—two clinical presentations the Dutch *explicitly disqualified* from their studies. As such, none of the Dutch findings are applicable to most of the youth seeking treatment today.

de Vries (2022) disputed only our assertion that the studies suffer from *high risk of bias* and therefore their findings of benefits are unreliable. She did not comment on our arguments that the research *failed to assess physical health risks* and *were not generalizable* to the majority of currently presenting cases. It is unclear if this silence indicates agreement or disagreement. Below, we address each of our points in greater detail, concluding with an additional observation about the overall lack of equipoise—genuine uncertainty about which treatment options are superior (London, 2017), which limits the utility of the Dutch research beyond describing a small-scale "innovative practice."

A. High risk of bias in the Dutch research

de Vries rejected our assertion that the Dutch findings suffer from a high risk of bias and insisted that we mistook the study protocol's careful process of establishing study eligibility for "bias." To clarify, we use the term "risk of bias" in a strict methodological sense. It refers to a systematic error, or deviation from the "truth" in study results (Boutron et al., 2022; Socialstyrelsen [National Board of Health and Welfare], 2022). Observational research conducted in the context of ongoing clinical care is often subject to risk of bias (Nguyen et al., 2021), which is one of the main reasons why rigorous clinical research using robust research designs must follow. In the case of the Dutch studies, we identified three major sources of bias, or systematic error, involving: (1) case selection; (2) measurement of outcomes, and (3) confounding.

1. Bias in case selection: Only the "best-case scenario" cases made it into the Dutch studies' "completers"

Because of an unusual case selection and reporting methodology, the Dutch studies inadvertently reported on only their best-case outcomes at each of the three phases of treatment (puberty blockers, cross-sex hormones, and surgery)—while failing to report the outcomes of the less positively affected, or even harmed, cases. de Vries disagreed with this assertion, continuing to insist that "participation was based on consecutive referral" (de Vries, 2022, p. 4).

Below, we present evidence that the claim of consecutive referral-based *prospective case selection* is not technically accurate. The actual case selection for the original sample of 70 puberty-blocked cases (de Vries et al., 2011) was *retrospective* and inadvertently biased toward including cases with favorable outcomes. The outcome reporting methodology in the second and final Dutch study (de Vries et al., 2014), which evaluated the final outcomes post-surgery, further biased the results toward reporting on the most favorable cases.

de Vries et al., 2011 ("puberty blocker" study). The 70 cases comprising the entire sample for the "puberty blocker" study (de Vries et al., 2011) were *retrospectively, non-randomly selected* from a larger group of consecutively referred 111 cases. According to both the original study, and de Vries' response to us, to participate in the "puberty blocker" study, a study subject already had to be starting the *next phase* of treatment with cross-sex hormones:

> Of the 196 consecutively referred adolescents…111 (those below age 16) had started puberty suppression… In the 2011 study we evaluated the first 70 of those 111 who were about to start with the next step of their treatment, affirming hormones, around the age of 16 years. (de Vries, 2022, p. 4)[2]

Using the start date of the *next phase* of treatment (cross-sex hormones) as the defining inclusion criterion for the study of the *prior phase* of the treatment (puberty blockers) introduced serious bias.

First, had any of the original 111 study subjects been harmed by puberty blockers or chosen to stop the treatment, they would never have advanced to the next phase, and thus, they had no chance of being included in the puberty blocker study, skewing the sample. *Second*, since the Dutch considered the puberty suppression phase both a treatment and a *diagnostic phase* (Cohen-Kettenis & van Goozen, 1998), the more complex cases may have remained in the puberty blocked phase longer. As de Vries' predecessors explained, subjects for whom the psychotherapist or parents had doubts, or where "the personal situation of the youngster" was more complicated, were delayed from starting cross-sex hormone treatment, which was the first stage the Dutch researchers considered to have an "irreversible" effect (Gooren & Delemarre-van de Waal, 1996, p. 11). This would further skew "the first 70 of those 111 who were about to start with the next step of their treatment, affirming hormones" (de Vries, 2022, p. 4)—the entire puberty blocker study sample—toward the most clinically straightforward and stable cases.

Third, such an unusual case selection methodology may have skewed the sample toward an older age than was stipulated by the protocol. Since to be eligible for the "puberty blocker" study, a subject had to have been deemed ready to start the next phase of cross-sex hormones, which *required a minimum age of 16* (accroding to the Dutch protocol version published in 2012, de Vries, 2012), all else being equal, older subjects had a greater chance of being included than younger ones. This may explain why the sample of 70 selected subjects was on average, age 15 when started on puberty blockers rather than age 12 as outlined by the protocol, which introduced another source of systematic error, by biasing the sample toward subjects with greater physical and cognitive maturity.

Given that the 2011 Dutch study's main goal was to evaluate the novel use of *puberty blockers* for gender dysphoria in a prospective cohort study (de Vries et al., 2011), the study should have enrolled, and reported the outcomes of, *all of the intent to treat* cases based on the date of eligibility to start *puberty suppression*—not cross-sex hormones.

It is notable that the only attempt to replicate the 2011 Dutch study results with more than a handful of cases took place in the UK but failed (Carmichael et al., 2021), with the conclusion of "no changes in psychological function" (p. 1). We suspect the key reason for this failure was the fact that the UK researchers truly *prospectively* selected "sequentially eligible" cases for treatment (Carmichael et al., 2021, p. 4) and as a result, ended with a diverse range of outcomes, including worsening of problems among female subjects during puberty blockade (Biggs, 2020). In contrast, the Dutch *retrospective* case selection methodology (misunderstood as prospective) inadvertently resulted in skewing the sample toward the best-case-scenario puberty-blocked cases. In our view, such case selection methodology invalidates the 2011 study conclusions of

psychological benefits of puberty suppression—or, as research methodologists would say, puts this finding at a "critical risk of bias."

de Vries et al, 2014 (post-surgery study). Skewing the sample toward the best-case scenario cases is even more apparent in the 2014 study, which reported on post-surgical outcomes and assessed the entire "gender-affirmative" treatment pathway (de Vries et al., 2014). The 70 participants who began the 2014 study, already biased toward more positive outcomes, shrank to 55. Fifteen subjects were dropped from the study and relabeled "nonparticipants." This subset, however, was not random, but instead heavily skewed toward subjects who experienced serious problems, including 3 who developed severe diabetes and obesity and 1 death following surgical complications. There is also considerable uncertainty about the outcomes of the 5 of 70 subjects (refusal, failure to return questionnaire, and dropping out of care) who, after several years of close contact with the research team, were unwilling to engage further:

> Nonparticipation (n = 15, 11 transwomen and 4 transmen) was attributable to not being 1 year postsurgical yet (n = 6), refusal (n = 2), failure to return questionnaires (n = 2), being medically not eligible *(e.g., uncontrolled diabetes, morbid obesity) for surgery (n = 3)*, dropping out of care (n =1), and *1 transfemale died after her vaginoplasty owing to a postsurgical necrotizing fasciitis* [emphasis added]. (de Vries et al., 2014, p. 697)

In her response, de Vries repeated the assertion that because a statistical comparison of the 15 "nonparticipants" to the 55 "participants" revealed no significant difference in their *pretreatment* baseline characteristics, "the results of the 2014 study can be generalized with substantial trust to the complete group of 70" (de Vries, 2022, pp. 4–5). We strongly disagree. The "participant" and "nonparticipant" cohorts are demonstrably different: while 100% of the 55 "participants" had successful gender reassignment according to the study reporting, at least 27% of the "nonparticipant" group (4/15: 1 death and 3 cases of diabetes) did not. Not only is a statistical analysis of such small subgroups massively underpowered to detect differences, ***no*** statistical analysis of *pretreatment* data suggesting "similarity" can negate the reality of the markedly different *post-treatment* outcomes in two groups. Nor is it clear why the research team made the unusual decision to stop the study early, before the remaining 6 participants had a chance to complete the 1-year post-surgical follow-up.

A note on the "missing Dutch study" on the effect of cross-sex hormones. The second and final Dutch study (de Vries et al., 2014) combined the cross-sex hormone and post-surgical treatment results into a single set of outcomes. This conflation may have made some sense at the time, as all the hormonally-treated patients were *required* to undergo surgery (removal of breasts, ovaries, uterus, penis, testes, and construction of a neovagina) by the Dutch protocol at the time. When surgery is not required, only 25–35% of transgender-identified adults appear to seek "gender-affirming" surgical procedures (Nolan et al., 2019). According to recently published data, this number is even smaller for youth: for every teen treated surgically, there are 15 treated *only* with cross-sex hormones (Respaut & Terhune, 2022). The inability of the Dutch research to elucidate the outcomes of cross-sex hormone treatments (separate from surgery) has been noted by NICE, which appropriately excluded the 2014 Dutch study from its systematic review of evidence (NICE, 2020b).

It is unknown whether the 4.3% of the sample (n = 3) that experienced obesity and diabetes sometime before the surgery was a result of the hormonal treatment; this rate appears to be double the expected rate for pediatric populations in the Netherlands at the time (Rotteveel et al., 2007; Schönbeck et al., 2011). Nor is it known if the cross-sex hormones contributed to the one subject who discontinued treatment due to other medical or psychological problems. Other research suggest that testosterone may actually *increase* dysphoria in female gender-dysphoric individuals (Olson-Kennedy, Warus, et al., 2018).

2. Bias in measurement of outcomes: The finding of "resolution of gender dysphoria" is invalid

The linchpin result of the Dutch studies is the reported *resolution of gender dysphoria*, as measured by the Utrecht Gender Dysphoria Scale (UGDS) (Steensma, Kreukels, et al., 2013). de

Vries agreed with us on this point: "the main finding remains the resolution of gender dysphoria" (de Vries, 2022, p. 3). According to the final Dutch study, the UGDS *gender dysphoria* scores plummeted, from a near-maximum score of 54 (maximum of 60) at baseline, to the near-minimum score of 16 (minimum of 12) after the final surgery (de Vries et al., 2014).

Rather than a true "resolution" of *gender dysphoria*, however, this spectacular drop was an artifact of switching the scale from "female" to "male" versions (and vice versa) before and after treatment, prompting a problematic *reversal* in the scoring. We argued that this fact alone invalidates the study's main conclusion of the resolution of gender dysphoria (Levine et al., 2022a). While de Vries conceded the use of the UGDS scale post-treatment was "not ideal" because "the UGDS was not…designed to be used after treatment," she asserted that it "does not imply that UGDS 'falsely' measured the improvement in GD [gender dysphoria]" (de Vries, 2022, p. 4). We think it is vitally important for the scientific community to recognize that the UGDS scale use was not merely "not ideal"—but that it *entirely invalidated* the Dutch study's main finding.

The following hypothetical scenario clearly demonstrates the problem. A severely gender dysphoric, cross-sex identified female patient is asked to answer two of the UGDS questions: "Every time someone treats me like a girl I feel hurt" and "Every time someone treats me like a boy I feel hurt" (Items 2 on the "female" and the "male" versions of the UGDS scale, respectively). It is likely that the patient would *strongly agree* with the first statement, and *strongly disagree* with the second. The first answer would lead to the score of "5" on the UGDS gender dysphoria scale, indicating the highest possible level of gender dysphoria. The second answer—which is effectively the same answer—would result in the score of "1" indicating the lowest possible gender dysphoria. This is because unlike the first question, which belongs to the "female" battery of questions, the second question belongs to the "male" battery of questions and effectively assumes the subject to be male—hence, the lack of distress of being associated with "maleness" receives the minimum "gender dysphoria" score.

If we now consider that only the "female" scale was used for gender dysphoric females at baseline but was then switched to the "male" scale after the final surgery (and vice-versa for male subjects), it becomes clear that the remarkable drop in "gender dysphoria" the UGDS scale registered after surgery entirely results from switching the scale. The *same* gender dysphoric individual, effectively answering the *same* question (albeit linguistically inverted), in the *same* way results in either the maximum or the minimum "gender dysphoria" score—depending on which sexed version of the scale was used. We reproduced both the "male" and the "female" versions of the UGDS scale in Table 1 so that others can easily observe how switching the scale "sex" version consistently leads to a "drop" of the gender dysphoria score, regardless of any treatment effect.

When defending the choice to reverse the UGDS scale (de Vries, 2022), de Vries pointed out—and we agree—that it would make no sense to ask postoperative natal males to answer a question such as "I dislike having erections" (Table 1, UGDS-M, item 11), since they no longer have penises. We empathize with the Dutch researchers' plight, as they found themselves without a valid tool to measure the construct of "gender dysphoria" after treatment. It is equally nonsensical, however, to ask natal males to answer questions such as, "I hate menstruating because it makes me feel like a girl" (Table 1, UGDS-F, item 10)—and it makes even less sense to report "resolution of gender dysphoria" because they don't "hate menstruating."

In her response, de Vries pointed to the validation research of the UGDS dysphoria scale (de Vries, 2022; Steensma, Kreukels, et al., 2013). To the best of our knowledge, this work has never appeared in a peer-reviewed publication. In our opinion, this UGDS validation research missed a key opportunity to identify the threat to validity of using the UGDS scale in post-gender reassignment context, which should have become apparent to the Dutch research team by 2013 when the validation paper was published. The greater community of international gender clinicians relying on the Dutch pioneering experience was not alerted to the need to find another instrument that can provide a valid pre-post "gender dysphoria" measure. Instead, the validation

Table 1. Utrecht Gender Dysphoria Scale, Adolescent Version (de Vries, Cohen-Kettenis, & Delemarre-van de Waal, 2006). Response categories are *agree completely, agree somewhat, neutral, disagree somewhat, disagree completely*.

UGDS-F (female) Response categories are: agree completely, agree somewhat, neutral, disagree somewhat, disagree completely. Items 1, 2, 4–6 and 10–12 are scored from 5 to 1; items 3 and 7–9 are scored from 1 to 5.	UGDS-M (male) Response categories are: agree completely, agree somewhat, neutral, disagree somewhat, disagree completely. Items are all scored from 5 to 1.
1. I prefer to behave like a boy.	1. My life would be meaningless if I would have to live as a boy.
2. Every time someone treats me like a girl I feel hurt.	2. Every time someone treats me like a boy I feel hurt.
3. I love to live as a girl.	3. I feel unhappy if someone calls me a boy.
4. I continuously want to be treated like a boy.	4. I feel unhappy because I have a male body.
5. A boy's life is more attractive for me than a girl's life.	5. The idea that I will always be a boy gives me a sinking feeling.
6. I feel unhappy because I have to behave like a girl.	6. I hate myself because I'm a boy.
7. Living as a girl is something positive for me.	7. I feel uncomfortable behaving like a boy, always and everywhere.
8. I enjoy seeing my naked body in the mirror.	8. Only as a girl my life would be worth living.
9. I like to behave sexually as a girl.	9. I dislike urinating in a standing position.
10. I hate menstruating because it makes me feel like a girl.	10. I am dissatisfied with my beard growth because it makes me look like a boy.
11. I hate having breasts.	11. I dislike having erections.
12. I wish I had been born as a boy.	12. It would be better not to live than to live as a boy.

research buttressed the problematic practice of using UGDS to measure the level of gender dysphoria after gender reassignment by stating: "From follow-up studies it was already known that gender dysphoria, as measured by the UGDS, disappeared post gender reassignment. These qualities make the instrument useful for clinical and research purposes" (Steensma, Kreukels, et al., 2013, p. 56). This statement is misleading, as the finding of the "disappearance" of gender dysphoria post-gender reassignment in the past "follow-up" research came from studies that also switched the sexed scale versions post-treatment, as Dr. de Vries pointed out in her response to us (de Vries, 2022).

Thus, in a spectacular display of circular reasoning, the scale validation research claimed that the follow-up research endorsed the use of the inverted UGDS scale version post gender reassignment, while the follow-up research defended this unusual practice by pointing to the validation research. de Vries doubled down on this circular reasoning in her response to our critique (de Vries, 2022):

> Levine et al. (2022) questions whether the improvement in gender dysphoria does then not stem from this switching, and not from the treatment? However, this seems turning the matter around. What the measure shows, the disappearance or resolution of gender dysphoria, is what the gender affirming treatment is aimed to resolve. (pp. 3–4)

At least three research groups noted the critical threat to the validity of the finding of "resolution of gender dysphoria" due to the switching of the scale (Biggs, 2022; McGuire et al., 2020; van de Grift et al., 2017). McGuire et al. (2020) explicitly stated, "Because the original UGDS is composed of two scales, it is impossible to determine if this is a real difference in gender dysphoria between groups or if this is an artifact of measurement error (p. 195).

The likely meaning of the "plummeting" gender dysphoria scores. What, if anything, did the "plummeting" gender dysphoria scores post scale-flipping signal, if not the "disappearance of gender dysphoria" claimed by the Dutch researchers? We posit that the UGDS scale can only measure the construct for which it was originally designed and validated to measure—the level of incongruence between natal sex and gender identity leading to the provision of the DSM diagnosis (Cohen-Kettenis & van Goozen, 1997; Iliadis et al., 2020; Steensma, Kreukels, et al., 2013). This is true whether the scale is used before or after treatment, and whether the "treatment" in question is "gender-affirmation" with hormones and surgeries, psychotherapy, or mere "watchful waiting," with the scale administered at various time points.

The fact that after gender reassignment, the UGDS scores were low on the opposite-sex scale indicates that the subjects would have scored high on the natal sex scale, which corresponds to a *persistence in transgender identity.* This is the only plausible interpretation of the "plummeting" UGDS scores that survives in the context of the scale questions and the linguistic and numerical gymnastics the scale underwent in the post-gender-reassignment context. The finding of persistence of transgender identity is not unexpected, especially since the Dutch researchers selected subjects with lifelong extreme cross-sex identification and follow-up was only 1.5 years post-surgery. What it does *not* mean is that the feeling of "incongruence" resolved. This point is underscored by the long-term follow-up data on male-to-female Dutch transitioners, presented at the WPATH 2022 Symposium by Dr. van der Meulen (Steensma et al., 2022). Nearly a quarter of the participants have felt that their bodies were still too masculine, and over half have experienced shame for the "operated vagina" and fearful their partner will find out their post-surgical status—despite registering low "gender dysphoria" UGDS scores (Steensma et al., 2022).

3. Bias from confounding: Psychotherapy was comingled with medical interventions

Although the Dutch research is frequently commended for having demonstrated "psychological improvements," an examination of the outcomes reveals that standard measures of psychological functioning such as anxiety, depression, anger, and global function showed very little clinically significant change after treatment (Levine et al., 2022a). de Vries acknowledged that a number of psychological measures showed no meaningful change, but insisted that the "more robust" measures, such as Child Behavior Check List (CBCL) and Youth Self Report (YSR), *did* show clinically relevant changes (de Vries, 2022, p. 3). She also noted that post-intervention, the sample of gender dysphoric youth in the Dutch research functioned at a similarly high level as their non-dysphoric peers, which was also an indicator of success. We have three observations about this response.

First, the impressive drop in the percentage of cases in the "clinical" range for CBCL and YSR (de Vries et al., 2014) was only apparent after *dichotomizing* these scales into the "clinical" (problematic) versus "non-clinical" ranges. In comparison, the sample's *average* post-intervention score changes on these scales were much more modest. For example, while the 2014 Dutch study points out that the "percent in the clinical range dropped from 30% to 7% on the YSR/ASR," which looks like an impressive reduction, the *average* t-scores dropped from 54.72 before treatment, to 48.53 after surgery (de Vries et al., 2014, p. 702). Both before and after t-scores were 60—typically interpreted as having no clinically significant symptoms (Achenbach & Rescorla, 2001). This suggests the reported improvements in CBCL and YSR came from relatively small score changes, which are of limited clinical significance, even if in the process the clinical threshold is crossed for some cases.

Second, while de Vries points to the *post-treatment* similarity in function of the gender-dysphoric group to the general population as evidence of treatment success, it is not known how different the groups were from the general population *pretreatment*. According to earlier research by Cohen-Kettenis and van Goozen (1997), which presumably utilized similar selection criteria, "when both pre- and posttest group means were compared with Dutch normative data, *all scores turned out to be within the average range* [emphasis added]" (p. 269). Smith et al. (2001) confirm this and explicitly state that both pretreatment and post-treatment, the group of gender dysphoric youth selected for the interventions were "normal functioning" as compared to their age peers in the Netherlands (Smith et al., 2001, p. 477). If the sample used in the two Dutch studies, which was recruited several years later but used the same careful case selection criteria, bears resemblance to the sample described by this earlier Dutch research, then the reported post-treatment similarities in psychological function between the "treated" group and the general population of peers should not be attributed to gender reassignment.

Third, and perhaps most relevant to this discussion, is the question of whether *any* of the reported changes in post-treatment psychological function scores, clinically significant or not, can be reasonably attributed to gender reassignment—or if these changes were influenced by confounding factors not accounted for in the research design. As noted by the authors of the

CBCL and YSR scales that de Vries says she favors, "improvement in scores from before to after services does not prove that the services were responsible for improvement. Other explanations are possible, such as (a) children's problems tend to decrease as they get older; (b) the people providing the data may report improvements because they believe that the services helped, and (c) the test-retest attenuation effect (a general tendency for people to report fewer problems at a second assessment)" (Achenbach & Rescorla, 2001, p. 183).

In addition to the general sources of confounding in uncontrolled studies relying on "before and after" measures, a vital source of confounding in the Dutch studies has been hiding in plain sight: All the subjects received psychotherapy at the same time they were undergoing gender reassignment. This comingling of interventions makes it impossible to determine which of the interventions "worked."

Psychotherapy was a key element in the Dutch protocol. Contrary to the now-common but erroneous assertion by the U.S. gender medicine establishment that psychotherapy for gender dysphoria is akin to "conversion" and should be avoided or even banned (Cantor, 2020), the Dutch studies reveal that psychotherapy was a key element of the protocol. According to the Dutch protocol, "[i]n cases involving confusion about gender feelings, psychotherapy and peer support can be helpful in *resolving the confusion and coming to self-acceptance* [emphasis added]" (de Vries, Cohen-Kettenis & Delemarre-van de Waal, 2006, p. 87). Not only was psychotherapy thought to be beneficial, but apparently it was a core part of the intervention: "…the adolescents were all regularly seen by one of the clinic's psychologists or psychiatrists. Psychological or social problems could thus be timely addressed" (de Vries et al., 2011, p. 2281). The researchers acknowledge that psychotherapy "…may have contributed to the psychological well-being of these gender dysphoric adolescents" (de Vries et al., 2011, p. 2281).

A discussion of the utility of psychotherapy to ameliorate gender dysphoria and related psychological distress is outside the scope of this article, other than to point out that the results of at least two studies suggest that psychological interventions are associated with improvements in two of the outcome domains—*gender dysphoria* (van de Grift et al., 2017) and *global function* (Costa et al., 2015)—absent any medical interventions.

B. Incompleteness of evidence regarding risks

Failure to consider the physical health risks of "gender-affirming" endocrine and surgical interventions is another methodological weakness of the Dutch studies. This omission is surprising since the Dutch team hypothesized that hormonal interventions might adversely impact bone and brain development several years before their seminal studies commenced (Delemarre-van de Waal & Cohen-Kettenis, 2006, p. 134). As discussed earlier, the Dutch studies did, however, report on the cases that were reclassified from "participants" to "non-participants," and listed the reasons for the nonparticipation, which revealed a possible 6–7% rate of associated adverse events.

Several studies since have confirmed likely adverse health effects of hormonal interventions, although their long-term impact on future health is not yet known. Research suggests that youth treated with puberty blockers develop problems with bone density accrual (Biggs, 2021; Nokoff et al., 2022) and that bone density may be impaired even after treatment with cross-sex hormones is initiated (Klink et al., 2015). Other research suggests heightened insulin resistance (Nokoff et al., 2021), elevated blood pressure, elevated triglycerides, and impaired liver function (Olson-Kennedy, Okonta, et al., 2018). Cross-sex hormone administration places adolescents in the medical category of early life indicators of future cardiovascular disease (Jacobs et al., 2022).

These adverse changes, already evident after a relatively short period of hormonal interventions, do not bode well for long-term health, since "gender-affirming" hormones are prescribed with the presumption of ongoing, lifelong treatment essential for maintaining a masculinized or feminized appearance. It is likely that other medical risks will emerge in the future. Patients and their families cannot make informed decisions about a treatment when the physical health

risks are assumed to be minimal and not reported, and only the potential psychological benefits are considered.

C. Poor generalizability/applicability to currently presenting cases

Given the dramatic change in the epidemiology of youth gender dysphoria which occurred after the studies were published (Levine et al., 2022a), the question of the applicability of the Dutch research to the current clinical dilemmas is one of the most important questions to interrogate in the field of pediatric gender medicine today.

Generalizability/applicability questions whether "available research evidence can be directly used to answer the health and healthcare question at hand" (Schünemann et al., 2022). We asserted and continue to assert that the Dutch studies are not applicable/generalizable to most gender dysphoric youth presenting today. This is evidenced by two facts: (1) the most common profile of youth seeking gender transition today is an adolescent with postpubertal emergence of a transgender identity and significant uncontrolled mental health comorbidities; (2) the Dutch researchers explicitly disqualified such patients from their studies because of their concern that the risks of early gender transition might outweigh the benefits.

1. Most of today's adolescents have postpubertal onset of trans identity and comorbid mental illness

Until about a decade ago, most patients seen by gender clinics were very young boys who wished to be girls and most of these children subsequently lost their cross-sex identification before reaching adulthood (Hembree et al., 2017; Ristori & Steensma, 2016; Singh et al., 2021). Today, the majority are female adolescents (de Graaf et al., 2018; Kaltiala-Heino et al., 2018; Zhang et al., 2021) with previously gender-normative childhoods whose trans identity emerged around or after puberty (Hutchinson et al., 2020; Zucker, 2019). Many suffer from significant preexisting mental illness such as depression and anxiety or neurocognitive challenges such as autism spectrum disorder (ASD) or attention deficit hyperactivity disorder (ADHD) (Becerra-Culqui et al., 2018; de Graaf et al., 2021; Kaltiala-Heino et al., 2015; Kozlowska et al., 2021; Strang et al., 2018; Thrower et al., 2020).

The presentation of adolescent-onset gender dysphoria is not entirely new—what's new is its scale. As with many trends, the change occurred "gradually, then suddenly." While there was evidence of it in the mid-2000s, around 2014–2015 the presentation of pediatric gender dysphoria in the Western world sharply shifted, from childhood-onset that skewed toward males, to adolescent-onset with a preponderance of females with mental health problems (Aitken et al., 2015; de Graaf et al., 2018). The Dutch researchers began their experiments with pediatric gender transition well before this demographic shift began to dominate clinical presentations of youth gender dysphoria.

Finland's national pediatric gender program was among the first to sound the alarm regarding the changing epidemiology of gender dysphoria presentation in youth. In 2015, they began observing that the youth presenting for treatment were primarily females who "do not fit the commonly accepted image of a gender dysphoric minor" (Kaltiala-Heino et al., 2015). The Finnish researchers saw a new pattern of "severe psychopathology preceding onset of gender dysphoria," with 75% already in treatment for other psychiatric issues when their gender dysphoria emerged. By 2019, the Finnish gender program was in full-alarm mode: "Research on adolescent onset gender dysphoria is scarce, and optimal treatment options have not been established… The reasons for the sudden increase in treatment-seeking due to adolescent onset gender dysphoria/transgender identification are not known" (Kaltiala-Heino & Lindberg, 2019, p. 62). This changing epidemiology was noted by other Nordic countries as well (Kaltiala, Bergman, et al., 2020).

The novel presentation of youth gender dysphoria was also reported by the largest pediatric gender clinic in the world at the time, the UK's GIDS/Tavistock (de Graaf et al., 2018). The now-famous graph of the GIDS data shows a trickle of gender dysphoric youth in years past

turning into a tidal wave by 2015, with a significant overrepresentation of teen girls. Between 2009 and 2016, the number of gender dysphoric females increased more than 70 times (de Graaf et al., 2018). The UK researchers concluded:

> The steep increase in birth-assigned females seeking help from gender services across the age range highlights an emerging phenomenon. It is important to follow birth-assigned females' trajectories, to better understand the changing clinical presentations in gender-diverse children and adolescents and to monitor the influence of social and cultural factors that impact on their psychological well-being. (de Graaf et al., 2018, p. 4)

The number of gender dysphoric youth referrals in the UK doubled again between 2020–2021 and 2021–2022 (NHS, 2022b).

While U.S. population-level data are hard to come by due to the country's decentralized and highly fragmented health care system, recent research shows that the number of gender dysphoric teens has also sharply risen in recent years, with a nearly 70% increase just between 2020 and 2021 (Respaut & Terhune, 2022). Combined with U.S. medical chart data samples, which show that the composition of the population changed "from predominantly transfeminine to…predominantly transmasculine in children and adolescents" (Zhang et al., 2021, p. 390) and that over 70% of gender dysphoric youth had been diagnosed with ASD, ADHD and other mental health problems *before* their diagnosis of gender dysphoria (Becerra-Culqui et al., 2018), it is apparent that the U.S. has not been immune to this remarkable epidemiologic trend that has engulfed youth in the Western world.

This now-ubiquitous presentation of gender dysphoria in troubled adolescents with previously gender-normative childhoods lacks a DSM-5-TR descriptor (American Psychiatric Association [APA], 2022), leaving clinicians to refer to it by many names, including *adolescent-onset gender dysphoria*; *postpuberty adolescent-onset transgender history*; and *rapid-onset gender dysphoria (ROGD).* The latter term was introduced by a U.S. researcher (Littman, 2018). Despite the controversy that Littman's hypotheses generated in the gender medicine establishment (Marchiano, 2018*),* her research withstood a second round of rigorous peer review (Littman, 2020). Subsequent detransitioner research lent further support to the ROGD hypothesis, with patients themselves reporting "that their gender dysphoria began during or after puberty and that mental health issues, trauma, peers, social media, online communities, and difficulty accepting themselves as lesbian, gay, or bisexual were related to their gender dysphoria and desire to transition" (Littman, 2021, p. 15). Even WPATH, which in 2018 strongly objected to Littman's research (WPATH, 2018), conceded in its 2022 "Standards of Care 8" that while no one has attempted to replicate Littman's research, it is apparent that "[f]or a select subgroup of young people, susceptibility to social influence impacting gender may be an important differential to consider" (Coleman et al., 2022, p. S45).

The novel phenomenon of high numbers of young people declaring a transgender identity for the first time in adolescence, often in the context of preexisting mental illness and/or trauma and social difficulties, has been described by several other mental health clinicians (Hutchinson et al., 2020; Schwartz, 2021; Zucker 2019). The only exception to the trend of mentally struggling adolescents presenting with gender dysphoria is the Amsterdam gender clinic itself, which has also seen an influx of teens and the preponderance of girls, but apparently without the mental health problems (Arnoldussen et al., 2020). Nonetheless, writing for the American journal *Pediatrics*, de Vries recognized the emergence of this new clinical phenomenon, noting that "gender identity development is diverse, and a new developmental pathway is proposed involving youth with postpuberty adolescent-onset transgender histories" (de Vries, 2020, p. 1) and noting that "some case histories illustrate the complexities that may be associated with later-presenting transgender adolescents and describe that some eventually detransition (de Vries, 2020, p. 2).

2. The Dutch studies disqualified cases most commonly presenting today: Adolescents with recent-onset gender dysphoria, nonbinary identities, or mental illness

From the outset in the late 1990s when the Dutch researchers first began to report on the results of youth gender transitions, they made it clear that their focus was exclusively on youth with

complete cross-sex identification "from toddlerhood onwards" (Cohen-Kettenis & van Goozen, 1998, p. 1). Furthermore, there was a strict requirement of psychological stability:

> First, they must have shown a *lifelong extreme and complete crossgender identity/role* [emphasis added]. Around puberty these feelings and behaviors must have become more rather than less pronounced. Second, they must be *psychologically stable* [emphasis added] (with the exception of depressed feelings, which often are a consequence of their living in the unwanted gender role) and function socially without problems (e.g., have a supportive family, do well at school). (Cohen-Kettenis & van Goozen, 1997, p. 265)

Of note, youth with non-binary identities, common today (Green et al., 2022), were *ineligible* for medical interventions according to the Dutch protocol, and instead needed psychotherapy: "adolescents… whose wish for sex reassignment seems to originate from factors other than a genuine and complete cross-gender identity are *served best by psychological interventions* [emphasis added] (de Vries et al., 2006, pp. 87–88).

Thus, the Dutch protocol explicitly *excluded* the characteristics of adolescents presenting to clinics in recent years—those whose trans-identities emerged around puberty; non-binary presentations without the wish for a complete cross-sex reassignment; or cases of gender dysphoria accompanied by significant uncontrolled mental illness. The high level of psychological functioning of the Dutch cohort *at baseline* serves as evidence that these selection criteria were indeed followed at the time (de Vries et al., 2011). The fact that "gender-affirming" interventions are now provided to the very segment that was explicitly excluded from the eligibility in the foundational studies is alarming.

D. Failure to consider alternatives (lack of research equipoise)

The Dutch researchers began their research into treatments of gender-dysphoric adolescents with the *foregone conclusion* that children who had life-long gender dysphoria and who continue to be cross-sex identified as adolescents would inevitably grow up to be transgender-identified adults. This assumption, based on "expert observations" from a handful of cases (O'Malley & Ayad, 2022; Cohen-Kettenis & van Goozen, 1997), has never been tested in rigorous comparative research. Further, the research team assumed that the only feasible treatment for these adolescents is early gender transition, and that psychotherapy alone is ineffective—also without testing this assumption through research. This violates the key requirement of equipoise in research—the principle that clinical investigators must approach research with genuine uncertainty regarding diagnostic, prevention, and treatment options—and allocate individuals to interventions in a manner that allows for generation of new knowledge (Freedman, 1987; London, 2017).

In fact, as de Vries' response to us emphasizes, the Dutch researchers continue to hold such firm belief into the beneficial nature of gender reassignment for youth, that they are far more concerned with the risk of "nontreatment" with hormones and surgery than they are with the possibility that the youth undergoing transition may not have needed such drastic interventions (de Vries, 2022, p. 3). However, some of the earlier research on the "non-treated" gender-variant and gender dysphoric adolescents challenges the assumptions of the permanence of trans identity in teens.

1. Non-treatment of "referred" adolescents with significant mental illness

Because of the careful case selection, the Dutch protocol rejected some youth from eligibility for gender reassignment due to serious "psychological or environmental problems" (Smith et al., 2001, p. 473). According to the study that followed the trajectories of these youth, the majority no longer wished to undergo gender transition once they reached *adulthood.*

Smith et al. (2001) reported that individuals rejected from gender reassignment in adolescence found noninvasive ways to deal with their gender dysphoria, and gender dysphoria significantly diminished. Upon follow-up 1–7 years later, only 20% of the rejected subjects (6/27) underwent gender reassignment as adults, while 80% refrained from it. Among those who remained medically untreated and participated in follow-up research, a remarkable 80% (11/14) "*did not feel*

any regrets about having refrained from SR [sex reassignment] or being rejected…." Only 7% (1 of 14) expressed strong regret (Smith et al., 2001, p. 477).

Data from the study by Smith et al. (2001) raise the possibility that the majority of those rejected from hormonal interventions not only were unharmed by waiting but benefited from "nontreatment" with gender reassignment in adolescence. Unlike the medically and surgically treated subjects, the "rejects" completed uninterrupted physical and psychological development, avoided sterility, maintained their sexual function, eliminated their risk of iatrogenic harm from surgery, and avoided the need for decades of dependence on cross-sex hormones. These cases also demonstrate that the assumption that "adolescents do not desist" was not true even at the time the Dutch team first introduced gender transitions of youth. It is even less true now, with research showing 10-30% rates of detransition among those who were trans-identified in adolescence and young adulthood (Boyd et al., 2022; Hall et al., 2021; Roberts et al., 2022). The long-term follow-up data on the Dutch adolescent transitioner cohort recently presented at the WPATH 2022 Symposim (Steensma et al., 2022) also suggest that the rate of cross-sex identification was not as stable as originally expected, with a sizable percentage reporting one or more instances of identity changes after treatment completion, especially among the individuals on the autistic spectrum (Steensma et al., 2022).

2. Non-treatment of "gender variant" youth in a community sample

Another study, also from the Netherlands, that took place before the practice of pediatric gender transition became widespread (Steensma, van der Ende, et al., 2013), also sheds light on what happens when childhood and adolescent gender-variance remains medically untreated. This large prospective longitudinal study based on a community sample (n = 879) found that about 6% of children (n = 51) ages 7–8 in a community sample were identified as "gender variant." At follow-up 24 years later, when the subjects were on average in their early 30s, *not a single individual* from the previously "gender-variant" subgroup of 51 children sought to undergo gender reassignment, despite the availability of these services.

There are three noteworthy observations in this study. *First*, the rate of "gender variance" of 6% reported in the community sample is remarkably similar to the current rate of transgender identification in U.S. youth of 2–9% (Johns et al., 2019; Kidd et al. 2021). *Second*, the gender-variant children were roughly 8–15 times more likely to grow up to be gay, lesbian, or bisexual adults compared to gender-normative youth. Gender variance is a common precursor to future homosexuality (Korte et al., 2008) and in fact in the Dutch studies, 97% of youth were gay, lesbian, or bisexual relative to their natal sex (de Vries et al., 2011). *Third*, only *one* of the 879 individuals in the sample underwent a male-to-female gender reassignment as an adult—and the individual had *not* been deemed "gender-variant" as a child (Steensma, van der Ende, et al., 2013, p. 2729). This challenges the current focus on medical interventions at increasingly younger ages.

The fact that none of the "gender variant" children in the sample sought gender reassignment as adults, when the study was published in 2013, merits scrutiny. These children would have been coming "of age" just a few years before the Dutch researchers conceived of the notion of *juvenile transsexual* and began to offer gender reassignment to adolescents. Thus, these children just missed the clinical shift in the Dutch practice—and perhaps not coincidentally, apparently all avoided the lifelong burden of living as a gender-reassigned individual.

The title of de Vries' commentary, *Ensuring Care for Transgender Adolescents Who Need It* (de Vries, 2022) prompts us to pose two questions. First, has the availability of the Dutch protocol itself created the "need?" Second, absent clear criteria to separate a young person's "wish" from a "need," will research rigor be required to demonstrate that the benefits outweigh the risks?

II. Newer research claiming benefits of youth gender transition is even more flawed

de Vries acknowledged that the Dutch research suffers from some limitations but insisted that newer research has sufficiently addressed these problems. She criticized us for not including a

review of newer studies that "consistently demonstrate improved or stable psychological functioning, body image, or treatment satisfaction varying from three months to up to two years from the initiation of treatment" (de Vries, 2022, p. 5). We are familiar with the seven studies de Vries mentions—as well as a number of other recent studies. What these studies "consistently demonstrate" is the art of *spin*—a well-documented problem in biomedical research where researchers "distort the interpretation of results and mislead readers so that results are viewed in a more favorable light" (Chiu et al., 2017). Due to length concerns, we discuss only three examples— Carmichael et al. (2021), Costa et al. (2015), and Tordoff et al. (2022). Most of the current research on the purported benefits of "gender-affirming care" suffers from similar limitations.

The UK study of puberty blockers by Carmichael et al. (2021), which attempted to replicate the Dutch puberty blocker study's findings of psychological improvements (de Vries et al., 2011), *failed to demonstrate psychological improvements*, conceding that its results are "in contrast to the Dutch study" (Carmichael et al., 2021, p. 19). The study found problems in bone mass density accrual among puberty-blocked youth. These problematic findings take on a decisively positive spin in the study conclusions, which refocus the reader on the positive "overall patient experience of changes on GnRHa treatment"; dismiss bone density problems as merely "consistent with suppression of growth"; and camouflage the failure to replicate the psychological benefits of puberty suppression by simply stating, "we identified no changes in psychological function" (Carmichael et al., 2021, p. 2). de Vries aided in the positive interpretation of the results by recasting the lack of improvement in psychological function following puberty suppression, as a *positive* finding of "stable psychological function" (de Vries 2022, p. 5)—yet it has never been demonstrated that psychological function of gender dysphoric adolescents with high baseline mental health function, as was required by the study criteria, would be expected to deteriorate absent intervention.

Spin also characterizes Costa et al. (2015), which compared psychosocial functioning of gender dysphoric youth who were puberty-suppressed to those who were delayed for medical treatment and received only psychotherapy. By the end of the 18-month study period, both groups ended up in the same psychosocial functional range using the Children's Global Assessment Scale (CGAS): 61–70 (out of 100 points), corresponding to "[s]ome difficulty in a single area, but generally functioning pretty well" (Shaffer, 1983). This study can hardly be cited as evidence of the superiority of the medical approach and in fact points to the viability of providing noninvasive therapy as an alternative to puberty suppression. Yet, the authors focus their abstract on the fact that the puberty-blocked group had higher function after puberty suppression than before, ignoring the fact that both the puberty-suppressed and the psychologically-treated only groups improved and there was no statistically-significant difference betwen the two by the end of the study period (Biggs, 2019). Questions regarding the extent to which improvements in selfreported psychological measures could be due to the placebo effect of puberty blockers have been recently raised (Clayton, 2022).

The spin of Tordoff et al. (2022) is dramatic. This study claimed that puberty blockers and "gender-affirming" hormones produced a 60% reduction in depression after only one year. However, this conclusion is in stark contrast to the raw data: at baseline, 59% of the yet-to-be treated patients had *moderate to severe depression*; by the end of the study at 12 months, 56% were still moderately to severely depressed, despite receiving hormone treatment (Supplementary material of e Table 3 Tordoff et al., 2022). This unchanged rate of depression became an "observed 60% lower odds of depression" via a methodology that *inferred* the "improvement" in the *treated cases* from the reported "worsening" in the *untreated cases*. Indeed, the untreated cases in the study had depression rates of 83% by the end of the study period (n = 6), compared to 56% of the treated cases (n = 57), seemingly supporting the conclusion that treatment with hormones alleviates depression.

However, by basing their conclusion about the relative success of the "treated" on the finding of lack of success among the "untreated" cases, the researchers failed to consider that

they lost an astounding 80% of their "untreated" cohort by the end of the study (28 of 35); in contrast, the "treated" cohort largely remained (84%) enrolled. The high dropout rate in "untreated" subjects makes intuitive sense: the study took place in a gender clinic setting, the primary purpose of which is provision of gender transition services. Youth whose distress was ameliorated without the use of hormones would have little reason to stay enrolled in the clinic and participate in the ongoing research. However, what this also suggests is that the highest functioning "untreated" youth dropped out of the study. Thus, the entire conclusion that because "untreated" cases faired so poorly on measures of depression, anxiety, or suicidality, it must be that hormones given to the "treated" cases "worked," is invalid. There are other problems in the study, including the fact that the use of psychiatric medications was not accounted for in the analysis. The university was aware of the problems with this research but chose to remain silent because the study's optimistic conclusions were so well received by national news media outlets (Rantz, 2022).

These examples demonstrate why we do not share de Vries' optimism that the newer studies conducted since the publication of the two seminal Dutch studies provide any additional confidence in, or support for, the practice of youth gender transitions. Most of the current research into the practice of pediatric transition continues in the context of gender clinic settings, which are actively providing gender transition to willing youth. Such low-quality observational research not only lacks the ability to control for the multiple sources of bias due to limitations in research design, but also is often led by clinicians with vested intellectual, professional, and financial conflicts of interest (Prasad, 2013).

III. Suggestions for future research

We were pleased to learn that de Vries has been awarded a substantial research grant to continue to study the effects of the Dutch protocol (Amsterdam UMC, 2022a). We welcome her decision to study the effects of the Dutch protocol on the novel cohort of youth whose trans identity only emerged in adolescence, as we agree that it is important to know "whether medical treatment is …useful for this group or whether there are too many risks… such as regret afterwards" (Amsterdam UMC, 2022b).

However, we think the time has come to reexamine the entire 25 years of Dutch experience using rigorous methodologies, to answer the critical questions about the full range of risks and benefits that the use of the Dutch protocol. We offer five suggestions relating to both past and future research:

1. Conduct comprehensive retrospective research

There have been over 6600 referrals to the Amsterdam gender clinic alone between 2000 and 2019 (Steensma et al., 2022), with likely additional referrals to the other Dutch gender clinics over the same time period, as well as new referrals since 2019. A retrospective chart review of these referred patients, supplemented by the data from the Dutch health and civil records registries Registers in The Netherlands (2022) could allow researchers to reexamine its quarter-century of experience of gender transition of youth and their outcomes in a way that is methodologically sound. The analysis should include outcomes of *all* patients diagnosed with gender dysphoria as children, adolescents, or young adults, rather than focusing only on those who chose to pursue medical interventions and explicitly agreed to participate in research. This retrospective review should seek to examine the outcomes of medical transition, psychotherapy, and no intervention. The effects of each step of the Dutch protocol should be disaggregated to gain a better understanding of the benefits and risks at each stage, and the results should be analyzed by natal sex and the age of gender dysphoria onset as validated by medical records.

2. Focus on comparative outcomes

The importance of *comparative* research to determine optimal treatments has been known since the 1990s (Guyatt, 1993). Comparing "before" and "after" psychological outcomes tends to overstate benefits due to number of factors, including "regression to the mean" (Knapp, 2016). Gender dysphoric youth often seek help at the peak of their distress. That many such "extreme" situations tend to naturally revert to a milder state even without an intervention is a well-recognized clinical and statistical phenomenon. While randomization is still the gold standard to reliably estimate treatment effects, when it is not possible (as is the case with retrospective research), researchers should consider utilizing quasi-experimental research designs (Harris et al., 2006). Recent post-hoc analysis of the effects of "gender-affirming" surgery, which utilized propensity-score matching to construct comparator groups, is an example of such analysis (Bränström & Pachankis, 2020c).

3. Track a full range of health outcomes utilizing objective measures whenever possible

The current exclusive focus on psychological and sexual functioning and self-reports is insufficient. Research should include a more objective evaluation of the effects of gender reassignment interventions on bone, brain, cardiovascular health, malignancies, and overall morbidity and all-cause mortality. As mentioned earlier, retrospective chart reviews of the referred patient cohorts, supplemented with relevant data from the Dutch health and civil records registries, should provide sufficient information to estimate the longer-term impact of hormonal and surgical interventions on morbidity and mortality, while also documenting the incidence of osteoporosis, cardiovascular disease, and cancer, as well as rates of mental illness and suicidality/suicide.

4. Pre-specify primary and secondary outcome measures and consistently track them

The primary outcomes of pediatric gender reassignment have been a moving target. In 1997, the Dutch researchers stated that the decision to start gender transition had as its goal to improve the "psychological problems of untreated adolescents" (Delemarre-van de Waal & Cohen-Kettenis, 2006, p. 132), since transitions undertaken in adulthood were already adequately relieving the feeling of gender incongruence itself. In her commentary, however, de Vries stated that psychological function may not the "best indicator for the benefits of such treatment" and that "measures that assess what makes life most worth living…" are most appropriate (de Vries, 2022, p. 3). Yet in a recent interview, she stated that the best indicator of treatment benefits is "satisfaction with care" (O'Malley & Ayad, 2022, 54:36). Primary outcome measures that serve as the rationale for the intervention must be clearly stated, justified, and consistently tracked.

If relief of "gender dysphoria" is still considered a primary outcome by the Dutch research team, a new measure of gender dysphoria that can be validated in both the pre- and the post-treatment settings is urgently needed, as the UGDS scale's use post-treatment is invalid. The updated UGDS-GS scale (McGuire et al., 2020) currently favored by de Vries (de Vries, 2022), appears to be a derivative of the earlier UGDS scale, and therefore may suffer from similar limitations when used in post-gender-reassignment settings.

5. Focus on long-term outcomes

Until recently, the long-term outcomes on the cohort of 70/55 cases have been an unanswered question. It was partially answered in a recent WPATH Symposium presentation by the Dutch team, comprised of presentations by Drs. de Rooy, Asseler, van der Meulen, van der Miesen, and Steensma (Steensma et al., 2022). As we look forward to seeing these preliminary findings elucidated in the upcoming peer-reviewed publications, we note several concerns.

First, it appears that the follow-up research combined the earlier-treated cohorts with the later-treated ones. We hope to see the outcomes of the 70/55 cases reported separately from other cases, so that the original cohort's outcomes can be quantified. *Second*, only half of the treated cases engaged in follow-up research (Bazelon, 2022; Steensma et al., 2022). This can bias the results, as individuals who experience more difficulties with their gender transition are less likely to engage with the physicians who treated them (Vandenbussche, 2022). Much follow-up research that reports positive outcomes relies on self-reported data compromised by high dropout rates (D'Angelo, 2018). In contrast, research that utilizes medical records and objective outcome measures shows much less optimistic outcomes (Dhejne et al., 2011; Bränström & Pachankis, 2020a, 2020b, 2020c). To mitigate the non-response bias, the Dutch research team should leverage chart data for all the referred patients, and report objective health outcomes for the *entire cohort* that was treated.

Third, we are concerned by the apparent dismissal of reproductive regret, which affected more than a quarter of the patients according to the data presented by Asseler, as merely a problem of the past when sterilizing surgery was a requirement (Steensma et al., 2022). The current treatment protocol of blocking puberty at Tanner stage 2 followed by cross-sexhormones, endorsed by the Endocrine Society (Hembree et al., 2017) and WPATH (Coleman et al., 2022), will most likely lead to chemical sterility, just as the prior surgical protocol led to permanent surgically-induced sterility. There are currently no effective, established methods to preserve fertility of individuals whose gametes have not matured (Rosenthal, 2021).

Fourth, the reported relationship difficulties reported by Asseler, with over 60% of individuals in their early to mid-30's still single also deserve serious consideration. The apparent sexual difficulties reported for male-to-female transitioners by van der Meulen (around 70% have problems with libido, have pain during sex, or have problems with achieving orgasm), combined with reproductive challenges, may be contributing to this outcome. *Fifth*, the team's preliminary optimistic conclusions that early puberty blockade did not worsen sexual function appears to be based on a problematic combining Tanner stages 2 and 3. The development of sexual organs and fertility is significantly more advanced in Tanner stage 3, compared to stage 2. Whether or not the high rate of sexual problems found in the transitioned population may be related to blocking puberty at Tanner stage 2 needs to be investigated.

These newly reported data underscore an urgent need to determine whether the benefits of medical interventions outweigh the now much better understood risks.

Concluding thoughts

The question, "Just because we can, should we?" is not unique to pediatric gender medicine. What makes this arena exceptional is the radical, irreversible nature of "gender-affirming" medical and surgical interventions desired by the exponentially growing numbers of youth in the Western world. The recent changes announced by WPATH SOC 8—specifically the removal of minimum age limits for medical and surgical treatments, and the elimination of the "distress" requirement by switching from DSM-5-TR to ICD-11 diagnostic criteria (Coleman et al., 2022; Robles García & Ayuso-Mateos, 2019; World Health Organization, 2019)—takes the field further in a truly extraordinary direction whereby *any desired body modification* desired by a child or a young person becomes automatically "medically necessary."

Another unique aspect of the gender medicine field is that a number of clinicians tasked with caring for gender-distressed have taken on the role of political campaigners—and in doing so, have traded wisdom and nuance for blunt activism (Kuper et al., 2022; McNamara et al., 2022). Their insistence that today's gender-dysphoric teens are tomorrow's transgender adults, and that their future happiness and mere survival hinges on early access to gender reassignment, is demonstrably false. While still reported as "rare" by the gender medicine establishment (Coleman et al., 2022; McNamara et al., 2022), the rate of medical detransition is already 10%-30% just a few years following transition (Boyd et al., 2022; Hall et al., 2021; Roberts et al., 2022). These

numbers are likely to rise in the future as regret historically has taken over a decade to materialize (Dhejne et al., 2014). Not all of those who detransitioned will consider themselves harmed, but many will—and a number already have (Vandenbussche, 2022; Littman, 2021).

When clinician-activists misuse the eminence of their institutions and medical societies to deny or obfuscate important facts about pediatric gender transition—that puberty blockers are prescribed to peri-pubertal children as young as 8–9; that mastectomies are commonly provided to teens; that the wave of detransition is rising and already far exceeds what's been historically recorded; and that no other pediatric intervention of similarly drastic nature has ever been delivered at scale based such low quality of evidence (McNamara et al., 2022)—they may succeed in scoring a political or legal "victory" in the short-term, but they also contribute to the longer-term erosion of public trust in the medical profession. They also inadvertently contribute to medical harm.

The scale of the potential harm can be fully appreciated if one considers that an astounding 1 in 10–20 middle school, high school, and college students in the West currently claim a transgender identity (ACHA, 2022; Johns et al., 2019; Kidd et al. 2021). Adolescent mental health in general is at an all-time low (Centers for Disease Control and Prevention [CDC], 2022). Lesbian, gay and bisexual youth and those on the autism spectrum (Bradley, 2022) are at particularly high risk of refracting their gender-non-conformity through the prism of transgender identity. Youth referrals for gender reassignment having risen already several thousand percent in the last decade, and nearly doubled between 2020/2021 and 2021/2022 (NHS, 2022b; Respaut & Terhune, 2022). If these young patients' sense of urgency is confused with certainty about their future happiness, while a flawed evidence base is mistaken for proven safety and effectiveness of youth gender reassignment, harm at scale will ensue.

As physicians are increasingly instructed to widely adopt "gender identity screening" of adolescents to "facilitate and increase…the delivery of gender-affirming" interventions (Lau et al., 2021, p. 1) and are misled about the (very low) quality of research, an analogy of the opioid epidemic powerfully emerges. The gender medicine field must reflect on the parallels between the pain as the "fifth vital sign," the misuse of research (Porter & Jick, 1980; Zhang, 2017), the pressure to meet patient demands, and the role of powerful special interests during the height of the opioid epidemic—and the trends in pediatric gender medicine today.

The field of gender medicine has a short time to self-correct before a growing number of authorities step in and impose guardrails to safeguard youth. Public health authorities in Finland, Sweden, and most recently England have already done just that, sharply deviating from the WPATH's poorly evidenced recommendations in "SOC 7" (Dahlen et al., 2021), with no apparent intention to follow the updated "SOC 8" either (COHERE (Council for Choices in Health Care), 2020; Socialstyrelsen [National Board of Health and Welfare], 2022; NHS, 2022a). NHS England's decision to close GIDS/Tavistock—the world's biggest pediatric gender clinic—and to place the care of gender-distressed youth in established clinical settings that "maintain a broad clinical perspective," provide "strong links to mental health services," and do not "exceptionalise gender identity issues," (Cass, 2022; NHS, 2022b) is a vote of no-confidence in the WPATH-endorsed "gender-affirming" approach that dominates the "gender clinic" model of care.

The American medical establishment appears to be taking a different approach. Rather than acknowledging the problems with the gender-affirmation model of care, there is an apparent effort underway to retrospectively redefine what "gender-affirmation" is. Originally defined as comprised of the provision of hormones and surgery to youth (Table 2, Rafferty, 2018), more recently gender affirmation has been positioned as merely "holistic care." The American Academy of Pediatrics recently made a surprising and welcome statement that hormones and surgery are not the preferred treatment for gender dysphoric youth, and that in fact "for the vast majority of children, it recommends the opposite" (Szilagyi, 2022). Whether this statement will be followed by earnest efforts to restrict the provision of highly invasive interventions to exceptional situations and to endorse non-invasive psychosocial interventions as first line of treatment—instead of inappropriately conflating psychotherapy for gender dysphoria with "conversion"—remains to be seen.

The former era of eminence-based, expert-opinion-led medicine, under which the innovative clinical practice of pediatric gender transition proliferated, has been replaced by a new standard, *evidence-based medicine*, which demands rigor in the research that underpins population-level treatment recommendations (Sackett et al., 1996; Zimerman, 2013). Our analysis of the Dutch protocol has been written with three goals in mind. *First*, we wanted to definitively refute the claims that the foundational Dutch research represents "solid prospective research" that provides reliable evidence of net benefits of youth gender transition. In fact, it is much better described as case series—one of the lowest levels of evidence available (Dekkers et al., 2012, Mathes & Pieper, 2017). *Second*, we aimed to demonstrate that the type of non-comparative, short-term research that the gender medicine establishment continues to pursue is incapable of generating reliable information. And *third and most importantly*, we wanted to remind the medical community that medicine is a double-edged sword capable of both much good and much harm. The burden of proof—demonstrating that a treatment does more good than harm—is *on those promoting the intervention*, not on those concerned about the harms. Until gender medicine commits to conducting high quality research capable of reliably demonstrating the preponderance of benefits over harms of these invasive interventions, we must be skeptical of the enthusiasm generated by headlines claiming that yet another "gender study" proved benefits of transitioning youth. This time-honored concern about risk/benefit ratio is a sobering reminder that the history of medicine is replete with examples of "cures" which turned out to far more harmful than the "disease."

Notes

1. de Vries also served as a peer-reviewer of our original paper, Levine et al. (2022a).
2. While not central to our argument, de Vries' claim that the selection of the 111 participants from the original 196 was based only on the researchers' interest in those age 16 and under is contradicted by the data. According to Table 1 in de Vries et al. (2011), there was at least one natal female participant who was 18.6 years old when the puberty blockers were initiated. Although selection criteria of the 111 from 196 may have introduced additional bias, we are most concerned with bias in the subsequent selection of 70 from the 111.

Acknowledgements

The authors would like to thank all of those who helpfully read earlier versions of this manuscript and contributed in various ways. This includes our colleagues as well as the formal peer-reviewers. We would like to especially thank Michael Biggs for his thoughtful suggestions.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms & profiles*. Burlington, VT: University of Vermont Research Center for Children, Youth, & Families.

Aitken, M., Steensma, T. D., Blanchard, R., VanderLaan, D. P., Wood, H., Fuentes, A., Spegg, C., Wasserman, L., Ames, M., Fitzsimmons, C. L., Leef, J. H., Lishak, V., Reim, E., Takagi, A., Vinik, J., Wreford, J., Cohen-Kettenis, P. T., de Vries, A. L. C., Kreukels, B. P. C., & Zucker, K. J. (2015). Evidence for an altered sex ratio in clinic-referred adolescents with gender dysphoria. *The Journal of Sexual Medicine*, *12*(3), 756–763. doi:10.1111/jsm.12817

American College Health Association (ACHA). (2022). National College Health Assessment III. Undergraduate Student Reference Group. Data Report. Spring 2022.

American Medical Association (AMA). (2022). *AMA reinforces opposition to restrictions on transgender medical care*. Retrieved November 15, 2022, from https://www.ama-assn.org/press-center/press-releases/ama-reinforces-opposition-restrictions-transgender-medical-care

American Psychiatric Association (APA). (2022). *Diagnostic and statistical manual of mental disorders* (5th ed., text rev.). doi:10.1176/appi.books.9780890425787

Amsterdam UMC. (2022a, July 1). *Three Vidi grants for Amsterdam Public Health researchers*. Retrieved September 3, 2022 from https://www.amsterdamumc.org/en/research/institutes/amsterdam-public-health/news/three-vidi-grants-for-amsterdam-public-health-researchers.htm

Amsterdam UMC. (2022b, July 7). *Three Vidi grants for Amsterdam Public Health researchers*. Retrieved September 3, 2022 from https://www.amsterdamumc.org/nl/vandaag/800.000-euro-voor-onderzoek-naar-transgenderzorg-jongeren.htm

Arnoldussen, M., Steensma, T. D., Popma, A., van der Miesen, A. I. R., Twisk, J. W. R., & de Vries, A. L. C. (2020). Re-evaluation of the Dutch approach: Are recently referred transgender youth different compared to earlier referrals? *European Child & Adolescent Psychiatry*, *29*(6), 803–811. doi:10.1007/s00787-019-01394-6

Balon R. (2022). Commentary on Levine et al.: Festina Lente (Rush Slowly). *Journal of Sex & Marital Therapy*, *48*(8), 775–778. doi:10.1080/0092623X.2022.2055686

Bazelon, E. (2022, June 15). The battle over gender therapy. The New York Times. https://www.nytimes.com/2022/06/15/magazine/gender-therapy.html

Becerra-Culqui, T. A., Liu, Y., Nash, R., Cromwell, L., Flanders, W. D., Getahun, D., Giammattei, S. V., Hunkeler, E. M., Lash, T. L., Millman, A., Quinn, V. P., Robinson, B., Roblin, D., Sandberg, D. E., Silverberg, M. J., Tangpricha, V., & Goodman, M. (2018). Mental health of transgender and gender nonconforming youth compared with their peers. *Pediatrics*, *141*(5), e20173845. doi:10.1542/peds.2017-3845\

Biggs, M. (2019). A Letter to the Editor Regarding the Original Article by Costa et al.: Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. *The Journal of Sexual Medicine*, *16*(12), 2043. doi:10.1016/j.jsxm.2019.09.002

Biggs, M. (2020). Gender dysphoria and psychological functioning in adolescents treated with GnRHa: Comparing Dutch and English prospective studies. *Archives of Sexual Behavior*, *49*(7), 2231–2236. doi:10.1007/s10508-020-01764-1

Biggs, M. (2021). Revisiting the effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria. *Journal of Pediatric Endocrinology and Metabolism*, *34*(7), 937–939. doi:10.1515/jpem-2021-0180

Biggs, M. (2022). The Dutch protocol for Juvenile transsexuals: Origins and evidence. *Journal of Sex & Marital Therapy*. Advance online publication. doi:10.1080/0092623X.2022.2121238

Boyd, I. L., Hackett, T., & Bewley, S. (2022). Care of transgender patients: A general practice quality improvement approach. *Healthcare*, *10*(1):121. doi:10.3390/healthcare10010121

Boutron I, Page MJ, Higgins JPT, Altman DG, Lundh A, & Hróbjartsson A. (2022). Chapter 7: Considering bias and conflicts of interest among the included studies. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, & Welch VA (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. London: Cochrane. (updated February 2022). Available from www.training.cochrane.org/handbook.

Bradley, S. J. (2022). Understanding vulnerability in girls and young women with high-functioning autism spectrum disorder. *Women*, *2*(1), 64–67. doi:10.3390/women2010007

Bränström, R., & Pachankis, J. E. (2020a). Reduction in mental health treatment utilization among transgender individuals after gender-affirming surgeries: A total population study. *American Journal of Psychiatry*, *177*(8), 727–734. doi:10.1176/appi.ajp.2019.19010080

Bränström, R., & Pachankis, J. E. (2020b). Correction to Bränström and Pachankis. (2020). *American Journal of Psychiatry*, *177*(8), 734–734. doi:https://doi.org/10.1176/appi.ajp.2020.1778correction

Bränström, R., & Pachankis, J. E. (2020c). Toward rigorous methodologies for strengthening causal inference in the association between gender-affirming care and transgender individuals' mental health: Response to letters. *American Journal of Psychiatry*, *177*(8), 769–772. doi:10.1176/appi.ajp.2020.20050599

Brierley, J., & Larcher, V. (2009). Compassionate and innovative treatments in children: A proposal for an ethical framework. *Archives of Disease in Childhood*, *94*(9), 651–654. doi:10.1136/adc.2008.155317

Brignardello-Peterson, R., & Wiercioch, W. (2022). *Effects of gender affirming therapies in people with gender dysphoria: Evaluation of the best available evidence*. https://ahca.myflorida.com/letkidsbekids/docs/AHCA_GAPMS_June_2022_Attachment_C.pdf

Cantor, J. M. (2020). Transgender and gender diverse children and adolescents: Fact-checking of AAP policy. *Journal of Sex & Marital Therapy*, *46*(4), 307–313. doi:10.1080/0092623X.2019.1698481

Centers for Disease Control and Prevention (CDC). (2022). *New CDC data illuminate youth mental health threats during the COVID-19 pandemic*. https://www.cdc.gov/media/releases/2022/p0331-youth-mental-health-covid-19.html

Carmichael, P., Butler, G., Masic, U., Cole, T. J., De Stavola, B. L., Davidson, S., Skageberg, E. M., Khadr, S., & Viner, R. M. (2021). Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK. *PLOS ONE*, *16*(2), e0243894. doi:10.1371/journal.pone.0243894

Cass, H. (2022). *Entry 8—Beyond the headlines*. Retrieved August 19, 2022, from https://cass.independent-review.uk/entry-8-beyond-the-headlines/

Chiu, K., Grundy, Q., & Bero, L. (2017). 'Spin' in published biomedical literature: A methodological systematic review. *PLOS Biology*, *15*(9), e2002173. doi:10.1371/journal.pbio.2002173

Clayton, A. (2022). Gender-affirming treatment of gender dysphoria in youth: A perfect storm environment for the placebo effect—the Implications for research and clinical practice. *Archives of Sexual Behavior. Online ahead of print*. doi:10.1007/s10508-022-02472-8

Cohen-Kettenis, P. T., & van Goozen, S. H. M. (1997). Sex reassignment of adolescent transsexuals: A follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *36*(2), 263–271. doi:10.1097/00004583-199702000-00017

Cohen-Kettenis, P. T., & van Goozen, S. H. M. (1998). Pubertal delay as an aid in diagnosis and treatment of a transsexual adolescent. *European Child & Adolescent Psychiatry*, *7*(4), 246–248. doi:10.1007/s007870050073

Coleman, E., Bockting, W., Botzer, M., Cohen-Kettenis, P., DeCuypere, G., Feldman, J., Fraser, L., Green, J., Knudson, G., Meyer, W. J., Monstrey, S., Adler, R. K., Brown, G. R., Devor, A. H., Ehrbar, R., Ettner, R., Eyler, E., Garofalo, R., Karasic, D. H., … Zucker, K. (2012). Standards of care for the health of transsexual, transgender, and gender-nonconforming people, Version 7. *International Journal of Transgenderism*, *13*(4), 165–232. doi:10.1080/15532739.2011.700873

Coleman, E., Radix, A. E., Bouman, W. P., Brown, G. R., de Vries, A. L. C., Deutsch, M. B., Ettner, R., Fraser, L., Goodman, M., Green, J., Hancock, A. B., Johnson, T. W., Karasic, D. H., Knudson, G. A., Leibowitz, S. F., Meyer-Bahlburg, H. F. L., Monstrey, S. J., Motmans, J., Nahata, L., … Arcelus, J. (2022). Standards of care for the health of transgender and gender diverse people, Version 8. *International Journal of Transgender Health*, *23*(sup1), S1–S259. doi:10.1080/26895269.2022.2100644

Costa, R., Dunsford, M., Skagerberg, E., Holt, V., Carmichael, P., & Colizzi, M. (2015). Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. *The Journal of Sexual Medicine*, *12*(11), 2206–2214. doi:10.1111/jsm.13034

COHERE (Council for Choices in Health Care). (2020). *Palveluvalikoimaneuvoston Suositus: Alaikäisten Sukupuoli-identiteetin Variaatioihin Liittyvän Dysforian Lääketieteelliset Hoitomenetelmät. [Recommendation of the Council for Choices in Health Care in Finland: Medical Treatment Methods for Dysphoria Related to Gender Variance in Minors.]*. https://segm.org/Finland_deviates_from_WPATH_prioritizing_psychotherapy_no_surgery_for_minors

Dahlen, S., Connolly, D., Arif, I., Junejo, M. H., Bewley, S., & Meads, C. (2021). International clinical practice guidelines for gender minority/trans people: Systematic review and quality assessment. *BMJ Open*, *11*(4), e048943. doi:10.1136/bmjopen-2021-048943

D'Angelo, R. (2018). Psychiatry's ethical involvement in gender-affirming care. *Australasian Psychiatry*, *26*(5), 460–463. doi:10.1177/1039856218775216

de Graaf, N. M., Giovanardi, G., Zitz, C., & Carmichael, P. (2018). Sex ratio in children and adolescents referred to the gender identity development service in the UK (2009–2016). *Archives of Sexual Behavior*, *47*(5), 1301–1304. doi:10.1007/s10508-018-1204-9

de Graaf, N. M., Huisman, B., Cohen-Kettenis, P. T., Twist, J., Hage, K., Carmichael, P., Kreukels, B. P. C., & Steensma, T. D. (2021). Psychological functioning in non-binary identifying adolescents and adults. *Journal of Sex & Marital Therapy*, *47*(8), 773–784. doi:10.1080/0092623X.2021.1950087

de Vries, A. L. C., & Cohen-Kettenis, P. T. (2012). Clinical management of gender dysphoria in children and adolescents: The Dutch approach. *Journal of Homosexuality*, *59*(3), 301–320. doi:10.1080/00918369.2012.653300

Dekkers, O. M., Egger, M., Altman, D. G., & Vandenbroucke, J. P. (2012). Distinguishing case series from cohort studies. *Annals of Internal Medicine*, *156*(1_Part_1), 37. doi:10.7326/0003-4819-156-1-201201030-00006

Delemarre-van de Waal, H. A., & Cohen-Kettenis, P. T. (2006). Clinical management of gender identity disorder in adolescents: A protocol on psychological and paediatric endocrinology aspects. *European Journal of Endocrinology*, *155*(suppl_1), S131–S137. doi:10.1530/eje.1.02231

de Vries, A. L. C. (2020). Challenges in timing puberty suppression for gender-nonconforming adolescents. *Pediatrics*, *146*(4), e2020010611. doi:10.1542/peds.2020-010611

de Vries, A. L. C. (2022). Ensuring care for transgender adolescents who need it: Response to 'Reconsidering Informed Consent for Trans-Identified Children, Adolescents and Young Adults'. *Journal of Sex & Marital Therapy*. Advance online publication. doi:10.1080/0092623X.2022.2084479

de Vries, A. L. C., Cohen-Kettenis, P. T., & Delemarre-van de Waal, H. (2006). Clinical Management of gender dysphoria in adolescents. *International Journal of Transgenderism*, *9*(3–4), 83–94. doi:10.1300/J485v09n03_04

de Vries, A. L. C., McGuire, J. K., Steensma, T. D., Wagenaar, E. C. F., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2014). Young adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*, *134*(4), 696–704. doi:10.1542/peds.2013-2958

de Vries, A. L. C., Steensma, T. D., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2011). Puberty suppression in adolescents with gender identity disorder: A prospective follow-up study. *The Journal of Sexual Medicine*, *8*(8), 2276–2283. doi:10.1111/j.1743-6109.2010.01943.x

Dhejne, C., Lichtenstein, P., Boman, M., Johansson, A. L. V., Långström, N., & Landén, M. (2011). Long-term follow-up of transsexual persons undergoing sex reassignment surgery: Cohort study in Sweden. *PLoS ONE*, *6*(2), e16885. doi:10.1371/journal.pone.0016885

Dhejne, C., Öberg, K., Arver, S., & Landén, M. (2014). An analysis of all applications for sex reassignment surgery in Sweden, 1960–2010: Prevalence, incidence, and regrets. *Archives of Sexual Behavior*, *43*(8), 1535–1545. doi:10.1007/s10508-014-0300-8

Drescher, J. (2022). Informed consent or scare tactics? A response to Levine et al.'s "Reconsidering Informed Consent for Trans-Identified Children, Adolescents, and Young Adults." *Journal of Sex & Marital Therapy*. Advance online publication. doi:10.1080/0092623X.2022.2080780

Drisko, J. W., & Friedman, A. (2019). Let's clearly distinguish evidence-based practice and empirically supported treatments. *Smith College Studies in Social Work*, *89*(3–4), 264–281. doi:10.1080/00377317.2019.1706316

Earl, J. (2019). Innovative practice, clinical research, and the ethical advancement of medicine. *The American Journal of Bioethics*, *19*(6), 7–18. doi:10.1080/15265161.2019.1602175

Freedman, B. (1987). Equipoise and the ethics of clinical research. *New England Journal of Medicine*, *317*(3), 141–145. doi:10.1056/NEJM198707163170304

Gooren, L., & Delemarre-van de Waal, H. (1996). The feasibility of endocrine interventions in Juvenile transsexuals. *Journal of Psychology & Human Sexuality*, *8*(4), 69–74. doi:10.1300/J056v08n04_05

Green, A. E., DeChants, J. P., Price, M. N., & Davis, C. K. (2022). Association of gender-affirming hormone therapy with depression, thoughts of suicide, and attempted suicide among transgender and nonbinary youth. *Journal of Adolescent Health*, *70*(4), 643–649. doi:10.1016/j.jadohealth.2021.10.036

Guyatt, G. H. (1993). Users' guides to the medical literature: II. How to use an article about therapy or prevention A. Are the results of the study valid? *JAMA*, *270*(21), 2598. doi:10.1001/jama.1993.03510210084032

Hall, R., Mitchell, L., & Sachdeva, J. (2021). Access to care and frequency of detransition among a cohort discharged by a UK national adult gender identity clinic: Retrospective case-note review. *BJPsych Open*, *7*(6), e184. doi:10.1192/bjo.2021.1022

Harris, A. D., McGregor, J. C., Perencevich, E. N., Furuno, J. P., Zhu, J., Peterson, D. E., & Finkelstein, J. (2006). The use and interpretation of quasi-experimental studies in medical informatics. *Journal of the American Medical Informatics Association*, *13*(1), 16–23. doi:10.1197/jamia.M1749

Health and Human Services (HHS). (2022). Nondiscrimination in health programs and activities—Proposed rule. 87 FR 47824, pp. 47824–47920. Docket Number HHS-OS-2022-0012. Docket RIN 0945-AA17 https://www.federalregister.gov/documents/2022/08/04/2022-16217/nondiscrimination-in-health-programs-and-activities

Hembree, W. C., Cohen-Kettenis, P. T., Gooren, L., Hannema, S. E., Meyer, W. J., Murad, M. H., Rosenthal, S. M., Safer, J. D., Tangpricha, V., & T'Sjoen, G. G. (2017). Endocrine treatment of gender-dysphoric/gender-incongruent persons: An endocrine society* clinical practice guideline. *The Journal of Clinical Endocrinology & Metabolism*, *102*(11), 3869–3903. doi:10.1210/jc.2017-01658

Herrera-Perez, D., Haslam, A., Crain, T., Gill, J., Livingston, C., Kaestner, V., Hayes, M., Morgan, D., Cifu, A. S., & Prasad, V. (2019). A comprehensive review of randomized clinical trials in three medical journals reveals 396 medical reversals. *ELife*, *8*, e45183. doi:10.7554/eLife.45183

Higgins JPT., Thomas J., Chandler J., Cumpston M., Li T, Page MJ., Welch VA (Eds). (2022). Cochrane Handbook for Systematic Reviews of Interventions *version 6.3*. London: Cochrane. (updated February 2022). Available from www.training.cochrane.org/handbook

Hutchinson, A., Midgen, M., & Spiliadis, A. (2020). In support of research into rapid-onset gender dysphoria. *Archives of Sexual Behavior*, *49*(1), 79–80. doi:10.1007/s10508-019-01517-9

Iliadis, S. I., Axfors, C., Friberg, A., Arinell, H., Beckman, U., Fazekas, A., Frisen, L., Sandström, L., Thelin, N., Wahlberg, J., Södersten, M., & Papadopoulos, F. C. (2020). Psychometric properties and concurrent validity of the Transgender Congruence Scale (TCS) in the Swedish setting. *Scientific Reports*, *10*(1), 18701. doi:10.1038/s41598-020-73663-3

Janssen, A. (2022). 10/28/22 Florida boards of medicine and osteopathic medicine joint rules/legislative committee rule workshop. https://thefloridachannel.org/videos/10-28-22-florida-boards-of-medicine-and-osteopathic-medicine-joint-rules-legislative-committee-rule-workshop/ Accessed November 12, 2022

Jacobs, D. R., Jr, Woo, J. G., Sinaiko, A. R., Daniels, S. R., Ikonen, J., Juonala, M., Kartiosuo, N., Lehtimäki, T., Magnussen, C. G., Viikari, J., Zhang, N., Bazzano, L. A., Burns, T. L., Prineas, R. J., Steinberger, J., Urbina, E. M., Venn, A. J., Raitakari, O. T., & Dwyer, T. (2022). Childhood cardiovascular risk factors and adult cardiovascular events. *The New England Journal of Medicine*, *386*(20), 1877–1888. doi:10.1056/NEJMoa2109191

Johns, M. M., Lowry, R., Andrzejewski, J., Barrios, L. C., Demissie, Z., McManus, T., Rasberry, C. N., Robin, L., & Underwood, J. M. (2019). Transgender identity and experiences of violence victimization, substance use, suicide risk, and sexual risk behaviors among high school students—19 States and Large Urban School Districts, 2017. *Morbidity and Mortality Weekly Report*, *68*(3), 67–71. doi:10.15585/mmwr.mm6803a3

Kaltiala, R., Bergman, H., Carmichael, P., de Graaf, N. M., Egebjerg Rischel, K., Frisén, L., Schorkopf, M., Suomalainen, L., & Waehre, A. (2020). Time trends in referrals to child and adolescent gender identity services: A study in four Nordic countries and in the UK. *Nordic Journal of Psychiatry*, *74*(1), 40–44. doi:10.1080/08039488.2019.1667429

Kaltiala-Heino, R., & Lindberg, N. (2019). Gender identities in adolescent population: Methodological issues and prevalence across age groups. *European Psychiatry*, *55*, 61–66. doi:10.1016/j.eurpsy.2018.09.003

Kaltiala-Heino, R., Bergman, H., Työläjärvi, M., & Frisen, L. (2018). Gender dysphoria in adolescence: Current perspectives. *Adolescent Health, Medicine and Therapeutics*, *9*, 31–41. doi:10.2147/AHMT.S135432

Kaltiala-Heino, R., Sumia, M., Työläjärvi, M., & Lindberg, N. (2015). Two years of gender identity service for minors: Overrepresentation of natal girls with severe problems in adolescent development. *Child and Adolescent Psychiatry and Mental Health*, *9*(1), 9. doi:10.1186/s13034-015-0042-y

Keith, K. (2022, July 27). HHS proposes revised ACA anti-discrimination rule. *Health Affairs Forefront*. Advance online publication. doi:10.1377/forefront.20220727.369815

Kidd, K. M., Sequeira, G. M., Douglas, C., Paglisotti, T., Inwards-Breland, D. J., Miller, E., & Coulter, R. W. S. (2021). Prevalence of gender-diverse youth in an urban school district. *Pediatrics*, *147*(6), e2020049823. doi:10.1542/peds.2020-049823

Klink, D., Caris, M., Heijboer, A., van Trotsenburg, M., & Rotteveel, J. (2015). Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. *The Journal of Clinical Endocrinology & Metabolism*, *100*(2), E270–E275. doi:10.1210/jc.2014-2439

Knapp, T. R. (2016). Why is the one-group pretest–posttest design still used? *Clinical Nursing Research*, *25*(5), 467–472. doi:10.1177/1054773816666280

Korte, A., Goecker, D., Krude, H., Lehmkuhl, U., Grüters-Kieslich, A., & Beier, K. M. (2008). Gender identity disorders in childhood and adolescence. *Deutsches Ärzteblatt International*, *105*(48), 834–841. doi:10.3238/arztebl.2008.0834

Kozlowska, K., Chudleigh, C., McClure, G., Maguire, A. M., & Ambler, G. R. (2021). Attachment Patterns in Children and Adolescents With Gender Dysphoria. *Frontiers in Psychology*, *11*, 582688. doi:10.3389/fpsyg.2020.582688

Kuper, L. E., Cooper, M. B., & Mooney, M. A. (2022). Supporting and advocating for transgender and gender diverse youth and their families within the sociopolitical context of widespread discriminatory legislation and policies. *Clinical Practice in Pediatric Psychology*, *10*(3), 336–345. doi:10.1037/cpp0000456

Lau, J. S., Kline-Simon, A., Sterling, S., Hojilla, J. C., & Hartman, L. (2021). Screening for gender identity in adolescent well visits: Is it feasible and acceptable? *Journal of Adolescent Health*, *68*(6), 1089–1095. doi:10.1016/j.jadohealth.2020.07.031

Levine, S. B., Abbruzzese, E., & Mason, J. W. (2022a). Reconsidering informed consent for trans-identified children, adolescents, and young adults. *Journal of Sex & Marital Therapy*, *48*(7), 706–727. doi:10.1080/0092623X.2022.2046221

Levine, S. B., Abbruzzese, E., & Mason, J. W. (2022b). What are we doing to these children? Response to Drescher, Clayton, and Balon Commentaries on Levine et al., 2022. *Journal of Sex & Marital Therapy*. Advance online publication. doi:10.1080/0092623X.2022.2136117

Littman, L. (2018). Parent reports of adolescents and young adults perceived to show signs of a rapid onset of gender dysphoria. *PLOS ONE*, *13*(8), e0202330. doi:10.1371/journal.pone.0202330

Littman, L. (2020). The use of methodologies in Littman (2018) is consistent with the use of methodologies in other studies contributing to the field of gender dysphoria research: Response to Restar (2019). *Archives of Sexual Behavior*, *49*(1), 67–77. doi:10.1007/s10508-020-01631-z

Littman, L. (2021). Individuals treated for gender dysphoria with medical and/or surgical transition who subsequently detransitioned: A survey of 100 detransitioners. *Archives of Sexual Behavior*, *50*(8), 3353–3369. doi:10.1007/s10508-021-02163-w

London, A. J. (2017). Equipoise in research: Integrating ethics and science in human research. *JAMA*, *317*(5), 525. doi:10.1001/jama.2017.0016

Marchiano, L. (2018, March 1). Transgenderism and the social construction of diagnosis. Quillette. https://quillette.com/2018/03/01/transgenderism-social-construction-diagnosis/

Mathes, T., & Pieper, D. (2017). Clarifying the distinction between case series and cohort studies in systematic reviews of comparative studies: Potential impact on body of evidence and workload. *BMC Medical Research Methodology*, *17*(1), 107. doi:10.1186/s12874-017-0391-8

McNamara, M. (2022). 10/28/22 Florida boards of medicine and osteopathic medicine joint rules/legislative committee rule workshop. https://thefloridachannel.org/videos/10-28-22-florida-boards-of-medicine-and-osteopathic-medicine-joint-rules-legislative-committee-rule-workshop/ Accessed November 12, 2022

McNamara, M., Lepore, C., & Alstott, A. (2022). Protecting transgender health and challenging science denialism in policy. *New England Journal of Medicine*, *387*(21), 1919–1921. doi:10.1056/NEJMp2213085

McGuire, J. K., Berg, D., Catalpa, J. M., Morrow, Q. J., Fish, J. N., Nic Rider, G., Steensma, T., Cohen-Kettenis, P. T., & Spencer, K. (2020). Utrecht gender dysphoria scale—Gender spectrum (UGDS-GS): Construct validity among transgender, nonbinary, and LGBQ samples. *International Journal of Transgender Health*, *21*(2), 194–208. doi:10.1080/26895269.2020.1723460

Nguyen, V. T., Engleton, M., Davison, M., Ravaud, P., Porcher, R., & Boutron, I. (2021). Risk of bias in observational studies using routinely collected data of comparative effectiveness research: A meta-research study. *BMC Medicine*, *19*(1), 279. doi:10.1186/s12916-021-02151-w

National Health Service (NHS). (2022a, October 20). *Interim service specification for specialist gender dysphoria services for children and young people—Public consultation*. https://www.engage.england.nhs.uk/specialised-commissioning/gender-dysphoria-services/

National Health Service (NHS). (2022b, July 28). *Regional model for gender care announced for children and young people*. http://tavistockandportman.nhs.uk/about-us/news/stories/regional-model-for-gender-care-announced-for-children-and-young-people/

National Institute for Health and Care Excellence (NICE). (2020a). *Evidence review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria*. https://cass.independent-review.uk/nice-evidence-reviews/

National Institute for Health and Care Excellence (NICE). (2020b). *Evidence review: Gender-affirming hormones for children and adolescents with gender dysphoria*. https://cass.independent-review.uk/nice-evidence-reviews/

Nokoff, N., Ma, N., Moreau, K., & Rothman, M. S. (2022). *Bone Mineral Density in Transgender Youth on Gender Affirming Therapies*. https://www.endocrine.org/news-and-advocacy/news-room/2022/longer-treatment-with-puberty-delaying-medication-leads-to-lower-bone-mineral-density

Nokoff, N. J., Scarbro, S. L., Moreau, K. L., Zeitler, P., Nadeau, K. J., Reirden, D., Juarez-Colunga, E., & Kelsey, M. M. (2021). Body composition and markers of cardiometabolic health in transgender youth on gonadotropin-releasing hormone agonists. *Transgender Health*, *6*(2), 111–119. doi:10.1089/trgh.2020.0029

Nolan, I. T., Kuhner, C. J., & Dy, G. W. (2019). Demographic and temporal trends in transgender identities and gender confirming surgery. *Translational Andrology and Urology*, *8*(3), 184–190. doi:10.21037/tau.2019.04.09

Olson-Kennedy, J., Okonta, V., Clark, L. F., & Belzer, M. (2018). Physiologic response to gender-affirming hormones among transgender youth. *Journal of Adolescent Health*, *62*(4), 397–401. doi:10.1016/j.jadohealth.2017.08.005

Olson-Kennedy, J., Warus, J., Okonta, V., Belzer, M., & Clark, L. F. (2018). Chest reconstruction and chest dysphoria in transmasculine minors and young adults: Comparisons of nonsurgical and postsurgical cohorts. *JAMA Pediatrics*, *172*(5), 431. doi:10.1001/jamapediatrics.2017.5440

O'Malley, S., & Ayad, S. (Hosts). (2022, March 7). Pioneers series: Where it all started. The Dutch Researchers Steensma & De Vries (No. 66) [Audio podcast episode]. Gender: A Wider Lens. https://gender-a-wider-lens.captivate.fm/episode/66-pioneers-series-where-it-all-started-the-dutch-researchers-steensma-de-vries

Pasternack, I., Söderström, I., Saijonkari, M., & Mäkelä, M. (2019). *Lääketieteelliset menetelmät sukupuolivariaatioihin liittyvän dysforian hoidossa. Systemaattinen katsaus. [Medical approached to treatment of dysphoria related to gender variations. A systematic review.]*. 106. https://app.box.com/s/y9u791np8v9gsunwgpr2kqn8swd9vdtx

Porter, J., & Jick, H. (1980). Addiction rare in patients treated with narcotics. *New England Journal of Medicine*, *302*(2), 123–123. doi:10.1056/NEJM198001103020221

Prasad, V. (2011). The frequency of medical reversal. *Archives of Internal Medicine*, *171*(18), 1675. doi:10.1001/archinternmed.2011.295

Prasad, V. (2013). Why randomized controlled trials are needed to accept new practices: 2 medical worldviews. *Mayo Clinic Proceedings*, *88*(10), 1046–1050. doi:10.1016/j.mayocp.2013.04.026

Prasad, V., & Ioannidis, J. P. (2014). Evidence-based de-implementation for contradicted, unproven, and aspiring healthcare practices. *Implementation Science*, *9*(1), 1748–5908. doi:10.1186/1748-5908-9-1

Rafferty, J. (2018). Ensuring comprehensive care and support for transgender and gender-diverse children and adolescents. *Pediatrics*, *142*(4), e20182162. doi:10.1542/peds.2018-2162

Rantz, J. (2022, August 23). Despite 'concerning' transgender study, UW kept quiet because of positive coverage. MyNorthwest.Com https://mynorthwest.com/3437497/rantz-uw-med-used-false-claims-to-push-gender-affirming-care-for-trans-teen/

Registers in The Netherlands. (2022). EIT Health Scandinavia, https://www.eithealth-scandinavia.eu/biobanksregisters/registers/netherlands/. Accessed 14 Dec.

Respaut, R., Terhune, C. (2022, October 6). Putting numbers on the rise in children seeking gender care. Retrieved October 14, 2022, from https://www.reuters.com/investigates/special-report/usa-transyouth-data/

Ristori, J., & Steensma, T. D. (2016). Gender dysphoria in childhood. *International Review of Psychiatry*, *28*(1), 13–20. doi:10.3109/09540261.2015.1115754

Roberts, C. M., Klein, D. A., Adirim, T. A., Schvey, N. A., & Hisle-Gorman, E. (2022). Continuation of gender-affirming hormones among transgender adolescents and adults. *The Journal of Clinical Endocrinology & Metabolism*, *107*(9), e3937–e3943. doi:10.1210/clinem/dgac251

Robles García, R., & Ayuso-Mateos, J. L. (2019). ICD-11 and the depathologisation of the transgender condition. *Revista de Psiquiatría y Salud Mental (English Edition)*, *12*(2), 65–67. doi:10.1016/j.rpsmen.2019.01.002

Rosenthal, S. M. (2021). Challenges in the care of transgender and gender-diverse youth: An endocrinologist's view. *Nature Reviews Endocrinology*, *17*(10), 581–591. doi:10.1038/s41574-021-00535-9

Rotteveel, J., Belksma, E. J., Renders, C. M., Hirasing, R. A., & Delemarre-Van de Waal, H. A. (2007). Type 2 diabetes in children in the Netherlands: The need for diagnostic protocols. *European Journal of Endocrinology*, *157*(2), 175–180. doi:10.1530/EJE-06-0754

Sackett, D. L., Rosenberg, W. M. C., Gray, J. A. M., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *BMJ*, *312*(7023), 71–72. doi:10.1136/bmj.312.7023.71

SBU (Swedish Agency for Health Technology Assessment and Assessment of Social Services). (2022). *Hormonbehandling vid könsdysfori—Barn och unga En systematisk översikt och utvärdering av medicinska aspekter [Hormone therapy at gender dysphoria—Children and young people A systematic review and evaluation of medical aspects]*. https://www.sbu.se/contentassets/ea4e698fa0c4449aaae964c5197cf940/hormonbehandling-vid-konsdysfori_barn-och-unga.pdf

Schönbeck, Y., Talma, H., van Dommelen, P., Bakker, B., Buitendijk, S. E., HiraSing, R. A., & van Buuren, S. (2011). Increase in prevalence of overweight in Dutch children and adolescents: A comparison of Nationwide Growth studies in 1980, 1997 and 2009. *PLoS ONE*, *6*(11), e27608. doi:10.1371/journal.pone.0027608

Schünemann HJ, Vist GE, Higgins JPT, Santesso N, Deeks JJ, Glasziou P, Akl EA, Guyatt GH. (2022). Chapter 15: Interpreting results and drawing conclusions. In: Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ ,Welch VA (Eds.), *Cochrane Handbook for Systematic Reviews of Interventions version 6.3*. London: Cochrane. (updated February 2022). Available from www.training.cochrane.org/handbook.

Schwartz, D. (2021). Clinical and ethical considerations in the treatment of gender dysphoric children and adolescents: When doing less is helping more. *Journal of Infant, Child, and Adolescent Psychotherapy*, *20*(4), 439–449. doi:10.1080/15289168.2021.1997344

Shaffer, D. (1983). A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry*, *40*(11), 1228. doi:10.1001/archpsyc.1983.01790100074010

Singh, D., Bradley, S. J., & Zucker, K. J. (2021). A follow-up study of boys with gender identity disorder. *Frontiers in Psychiatry*, *12*, 632784. doi:10.3389/fpsyt.2021.632784

Smith, Y. L. S., Van Goozen, S. H. M., & Cohen-Kettenis, P. T. (2001). Adolescents with gender identity disorder who were accepted or rejected for sex reassignment surgery: A prospective follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *40*(4), 472–481. doi:10.1097/00004583-200104000-00017

Steensma, T. D., de Rooy, F. B. B., van der Meulen, I. S., Asseler, J. D., & van der Miesen, A. I. R. (2022, September 16–20). *Transgender Care Over the Years: First Long-Term Follow-Up Studies and Exploration of Sex Ratio in the Amsterdam Child and Adolescent Gender Clinic* [Conference presentation]. World Professional Association for Transgender Health Symposium, Montreal, QC, Canada.

Steensma, T. D., van der Ende, J., Verhulst, F. C., & Cohen-Kettenis, P. T. (2013). Gender variance in childhood and sexual orientation in adulthood: A prospective study. *The Journal of Sexual Medicine*, *10*(11), 2723–2733. doi:10.1111/j.1743-6109.2012.02701.x

Steensma, T. D., Kreukels, B. P., Jurgensen, M., Thyen, U., de Vries, A. L., Cohen-Kettenis, P. T. (2013). The utrecht gender dysphoria scale: A validation study. In: Steensma TD (Ed.), *From gender variance to gender dysphoria: Psychosexual development of gender atypical children and adolescents* (pp. 41–56). Amsterdam: Vrije Universiteit. https://research.vu.nl/ws/portalfiles/portal/42117766/table±of±contents.pdf

Strang, J. F., Meagher, H., Kenworthy, L., de Vries, A. L. C., Menvielle, E., Leibowitz, S., Janssen, A., Cohen-Kettenis, P., Shumer, D. E., Edwards-Leeper, L., Pleak, R. R., Spack, N., Karasic, D. H., Schreier, H., Balleur, A., Tishelman, A., Ehrensaft, D., Rodnan, L., Kuschner, E. S., … Anthony, L. G. (2018). Initial clinical guidelines for co-occurring autism spectrum disorder and gender dysphoria or incongruence in adolescents. *Journal of Clinical Child and Adolescent Psychology*, *47*(1), 105–115. doi:10.1080/15374416.2016.1228462

Socialstyrelsen [National Board of Health and Welfare]. (2022). *Care of children and adolescents with gender dysphoria – Summary*. Retrieved July 22, 2022 from https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/kunskapsstod/2022-3-7799.pdf

Szilagyi, M. (2022, August 21). Academy of pediatrics responds on trans treatment for kids. Wall Street Journal. https://www.wsj.com/articles/trans-gender-pediatric-aap-kids-children-care-surgery-affirm-treatment-11660942086

Thrower, E., Bretherton, I., Pang, K. C., Zajac, J. D., & Cheung, A. S. (2020). Prevalence of autism spectrum disorder and attention-deficit hyperactivity disorder amongst individuals with gender dysphoria: A systematic review. *Journal of Autism and Developmental Disorders*, *50*(3), 695–706. doi:10.1007/s10803-019-04298-1

Tordoff, D. M., Wanta, J. W., Collin, A., Stepney, C., Inwards-Breland, D. J., & Ahrens, K. (2022). Mental health outcomes in transgender and nonbinary youths receiving gender-affirming care. *JAMA Network Open*, *5*(2), e220978. doi:10.1001/jamanetworkopen.2022.0978

Turban, J. (2022, March 1). Opinion | Texas officials are spreading blatant falsehoods about medical care for transgender kids. Washington Post. Retrieved November 15, 2022, from https://www.washingtonpost.com/opinions/2022/03/01/texas-ken-paxton-greg-abbott-misinformation-transgender-medical-care/

van de Grift, T. C., Elaut, E., Cerwenka, S. C., Cohen-Kettenis, P. T., De Cuypere, G., Richter-Appelt, H., & Kreukels, B. P. C. (2017). Effects of medical interventions on gender dysphoria and body image: A follow-up study. *Psychosomatic Medicine*, *79*(7), 815–823. doi:10.1097/PSY.0000000000000465

Vandenbussche, E. (2022). Detransition-related needs and support: A cross-sectional online survey. *Journal of Homosexuality*, *69*(9), 1602–1620. doi:10.1080/00918369.2021.1919479

World Health Organization. (2019). International statistical classification of diseases and related health problems (11th ed.). https://icd.who.int/

World Professional Association for Transgender Health (WPATH). (2018). WPATH position on "Rapid-Onset Gender Dysphoria (ROGD)". Retrieved July 11, 2022, from https://www.wpath.org/media/cms/Documents/Public%20Policies/2018/9_Sept/WPATH%20Position%20on%20Rapid-Onset%20Gender%20Dysphoria_9-4-2018.pdf

Zhang, S. (2017, June 2). *The one-paragraph letter from 1980 that fueled the opioid crisis*. The Atlantic. https://www.theatlantic.com/health/archive/2017/06/nejm-letter-opioids/528840/

Zhang, Q., Rechler, W., Bradlyn, A., Flanders, W. D., Getahun, D., Lash, T. L., McCracken, C., Nash, R., Panagiotakopoulos, L., Roblin, D., Sandberg, D. E., Silverberg, M. J., Tangpricha, V., Vupputuri, S., & Goodman, M. (2021). Changes in size and demographic composition of transgender and gender non-binary population receiving care at integrated health systems. *Endocrine Practice*, *27*(5), 390–395. doi:10.1016/j.eprac.2020.11.016

Zimerman, A. (2013). Evidence-based medicine: A short history of a modern medical movement. *AMA Journal of Ethics*, *15*(1), 71–76. doi:10.1001/virtualmentor.2013.15.1.mhst1-1301

Zucker, K. J. (2019). Adolescents with gender dysphoria: Reflections on some contemporary clinical and research issues. *Archives of Sexual Behavior*, *48*(7), 1983–1992. doi:10.1007/s10508-019-01518-8

Human Reproduction, Vol.36, No.5, pp. 1427–1435, 2021
Advance Access Publication on January 17, 2021 doi:10.1093/humrep/deaa357

human reproduction

ORIGINAL ARTICLE ***Reproductive epidemiology***

Early initiation of anti-androgen treatment is associated with increased probability of spontaneous conception leading to childbirth in women with polycystic ovary syndrome: a population-based multiregistry cohort study in Sweden

E. Elenis[1], E. Desroziers[2], S. Persson[1], I. Sundström Poromaa[1], and R.E. Campbell[2,*]

[1]Department of Women's and Children's Health, Uppsala University, Uppsala 751 85, Sweden [2]Centre for Neuroendocrinology & Department of Physiology, School of Biomedical Sciences, Otago University, Dunedin 9054, New Zealand

Submitted on July 16, 2020; resubmitted on November 3, 2020; editorial decision on November 17, 2020

STUDY QUESTION: Is anti-androgen treatment during adolescence associated with an improved probability of spontaneous conception leading to childbirth in women with polycystic ovary syndrome (PCOS)?

SUMMARY ANSWER: Early initiation of anti-androgen treatment is associated with an increased probability of childbirth after spontaneous conception among women with PCOS.

WHAT IS KNOWN ALREADY: PCOS is the most common endocrinopathy affecting women of reproductive age. Hyperandrogenism and menstrual irregularities associated with PCOS typically emerge in early adolescence. Previous work indicates that diagnosis at an earlier age (<25 years) is associated with higher fecundity compared to a later diagnosis.

STUDY DESIGN, SIZE, DURATION: This population-based study utilized five linked Swedish national registries. A total of 15 106 women with PCOS and 73 786 control women were included. Women were followed from when they turned 18 years of age until the end of 2015, leading to a maximum follow-up of 10 years. First childbirth after spontaneous conception was the main outcome, as identified from the Medical Birth Registry.

PARTICIPANTS/MATERIALS, SETTING, METHODS: Participants included all women born between 1987 and 1996 with a diagnosis of PCOS in the Swedish Patient Registry and randomly selected non-PCOS controls (ratio 1:5). Information on anti-androgenic treatment was retrieved from the Swedish Prescribed Drug Registry with the use of Anatomic Therapeutic Chemical (ATC) codes. Women with PCOS who were not treated with any anti-androgenic medication were regarded as normo-androgenic, while those treated were regarded as hyperandrogenic. Women were further classified as being mildly hyperandrogenic if they received anti-androgenic combined oral contraceptive (aaCOC) monotherapy, or severely hyperandrogenic if they received other anti-androgens with or without aaCOCs. Early and late users comprised women with PCOS who started anti-androgenic treatment initiated either during adolescence ($\leq$ 18 years of age) or after adolescence (>18 years), respectively. The probability of first childbirth after spontaneous conception was analyzed with the use of Kaplan–Meier hazard curve. The fecundity rate (FR) and 95% confidence interval for the time to first childbirth that were conceived spontaneously were calculated using Cox proportional hazards regression models, with adjustment for obesity, birth year, country of birth and education level.

MAIN RESULTS AND THE ROLE OF CHANCE: The probability of childbirth after spontaneous conception in the PCOS group compared to non-PCOS controls was 11% lower among normo-androgenic (adjusted FR 0.68 (95% CI 0.64–0.72)), and 40% lower among hyperandrogenic women with PCOS (adjusted FR 0.53 (95% CI 0.50–0.57)). FR was lowest among severely hyperandrogenic women with

PCOS compared to normo-androgenic women with PCOS (adjusted FR 0.60 (95% CI 0.52–0.69)), followed by mildly hyperandrogenic women with PCOS (adjusted FR 0.84 (95% CI 0.77–0.93)). Compared to early anti-androgenic treatment users, late users exhibited a lower probability of childbirth after spontaneous conception (adjusted FR 0.79 (95% CI 0.68–0.92)).

LIMITATIONS, REASONS FOR CAUTION: We lacked direct information on the intention to conceive and the androgenic biochemical status of the PCOS participants, applying instead the use of anti-androgenic medications as a proxy of hyperandrogenism. The duration of anti-androgenic treatment utilized is not known, only the age at prescription. Results are not adjusted for BMI, but for obesity diagnosis. The period of follow-up (10 years) was restricted by the need to include only those women for whom data were available on the dispensing of medications during adolescence (born between 1987 and 1996). Women with PCOS who did not seek medical assistance might have been incorrectly classified as not having the disease. Such misclassification would lead to an underestimation of the true association between PCOS and outcomes.

WIDER IMPLICATIONS OF THE FINDINGS: Early initiation of anti-androgen treatment is associated with better spontaneous fertility rate. These findings support the need for future interventional randomized prospective studies investigating critical windows of anti-androgen treatment.

STUDY FUNDING/COMPETING INTEREST(S): This study was funded by the Health Research Council of New Zealand (18-671), the Swedish Society of Medicine and the Uppsala University Hospital. Evangelia Elenis has, over the past year, received lecture fee from Gedeon Richter outside the submitted work. Inger Sundström Poromaa has, over the past 3 years, received compensation as a consultant and lecturer for Bayer Schering Pharma, MSD, Gedeon Richter, Peptonics and Lundbeck A/S. The other authors declare no competing interests.

TRIAL REGISTRATION NUMBER: N/A

Key words: PCOS / fertility / childbirth / anti-androgen / adolescence

Introduction

Polycystic ovary syndrome (PCOS) is an endocrinopathy affecting women of reproductive age with a reported prevalence ranging from 5 to 25% (March *et al.*, 2010; Rosenfield and Ehrmann, 2016; Wolf *et al.*, 2018). PCOS is characterized by clinical or biochemical hyperandrogenism, menstrual irregularities and ultrasonographic polycystic ovarian morphology (Rosenfield and Ehrmann, 2016). Symptoms typically emerge during early adolescence (Driscoll, 2003; Ryan *et al.*, 2018) and may persist into adulthood. The common denominator for PCOS development appears to be ovarian and/or adrenal hyperandrogenism in synergy with tissue-selective insulin-resistant hyperinsulinism (Ibáñez *et al.*, 2017; Witchel *et al.*, 2019). The disorder is multifactorial and heterogeneous, implicating both intrauterine and postnatal environmental factors, as well as endocrinological, genetic and epigenetic factors (Rosenfield and Ehrmann, 2016). PCOS pathogenesis likely results from the combination of a prenatal predisposing factor (referred to as a 'first hit') with an activating postnatal factor (referred to as the 'second hit') (Rosenfield, 2020). For example, genetically susceptible girls or those exposed to androgen excess *in utero* develop hyperandrogenism prepubertally through hyperactivation of their hypothalamic–pituitary–ovarian (HPO) axis; that, in addition to the normal physiological or obesity-related hyperinsulinism during adolescence, potentiates the hyperandrogenic state and accelerates the syndrome's clinical manifestations and/or aggravates the syndrome's clinical course (Bremer, 2010). A more recent evolution of this idea suggests that a mismatch between prenatal and postnatal weight gain, resulting in greater hepatovisceral fat, drives accelerated body growth and maturation, which in turn establishes persistent PCOS features (de Zegher *et al.*, 2018).

In population-based studies (Koivunen *et al.*, 2008; West *et al.*, 2014; Persson *et al.*, 2019), we and others have previously demonstrated that women with PCOS, especially those with obesity, need a longer time to achieve childbirth and give birth to a lower number of children compared to non-PCOS counterparts. A novel finding was the fact that PCOS diagnosis at an earlier age (<25 years) was associated with higher fecundity rate (FR) compared to a later diagnosis (Persson *et al.*, 2019). Since symptoms appear to be progressive in women with PCOS, timely interventions that improve hyperandrogenism, either directly or indirectly through lowering insulin levels, have been recommended (Bremer, 2010). Therefore, whether specific interventions, such as pharmacological treatment during a specific therapeutic window, i.e. during adolescence, can decrease androgen actions and mitigate the future adverse effects of PCOS remains unknown.

Clinical and animal-based evidence indicates that long-term anti-androgen therapy can restore impaired reproductive function. Long-term AR blockade is associated with improved testosterone levels and ovulatory function in adult women with PCOS (De Leo *et al.*, 1998; Paradisi *et al.*, 2013), and a restoration of normal steroid hormone feedback to the reproductive axis (Eagleson *et al.*, 2000). In addition, in prenatally androgenized mice that model PCOS in adulthood (Sullivan and Moenter, 2004; Moore *et al.*, 2013; Moore *et al.*, 2015; Silva *et al.*, 2018), anti-androgen therapy restores estrous cyclicity (Sullivan and Moenter, 2004; Silva *et al.*, 2018). In addition, continuous androgen blockade from an 'adolescent' period following puberty is associated with improved ovarian morphology and a reversal of brain wiring changes induced by prenatal androgen exposure (Silva *et al.*, 2018).

Therefore, our study hypothesizes that the probability of childbirth after spontaneous conception among PCOS women improves if preceded by anti-androgen therapy during adolescence. The aim of the current study was therefore to explore whether treatment with anti-androgen medications initiated during adolescence is associated with a higher probability of childbirth after spontaneous conception in women with PCOS.

Materials and methods

Ethical approval

The study has been approved by the Regional Ethical Review Board in Uppsala, Sweden (Diary number 2017/309). The need for written or oral informed consent for the participating women in our study was weaved since all data received from the Swedish registries were anonymized.

Study design

The current study is part of a larger population-based project performed in Sweden on 45 395 women with PCOS and 217 049 non-PCOS controls. The study design has been previously presented by Persson *et al.* (2019). In summary, the data were assembled after linkage of five Swedish national registries, by utilizing the unique personal identification number that individuals are assigned at birth or immigration (Ludvigsson *et al.*, 2009). National registries comprise prospectively collected information from all inhabitants residing in the country and are maintained by Swedish government agencies such as the National Board of Health and Welfare and Statistics Sweden. The provided data in the present study arise from the Patient Registry, the Swedish Prescribed Drug Registry, the Medical Birth Registry, the Education Registry and the Total Population Registry (Ludvigsson *et al.*, 2009).

The Swedish Patient Registry comprises nationwide information on visiting dates and given diagnoses on both psychiatric and somatic care recorded during inpatient and outpatient visits. The visits include visits to gynecologists or fertility specialists. After 1997, diagnoses were classified according to the ICD-10 (International Statistical Classification of Diseases and Related Health Problems, version 10) (Ludvigsson *et al.*, 2011).

The Swedish Prescribed Drug Registry contains information on Anatomic Therapeutic Chemical (ATC) classification codes for prescribed and dispensed drugs, substances, brand names, formulations and daily dosages, together with the date of dispensing since 2005.

The Swedish Medical Birth Registry contains information on prenatal, delivery and neonatal care covering practically all births in Sweden since it was established in 1973. Information recorded includes prospectively collected demographic data, such as maternal age, reproductive history and assisted reproduction, and complications during pregnancy, delivery and the neonatal period.

The Swedish Education Registry, founded in 1985, contains data on demographics and educational attainment of the population.

The Total Population Registry, founded in 1968, contains data on life events including birth, death, place of residence and country of birth. It allows for identification of general population controls and estimation of follow-up time.

Exposure

We defined PCOS as presence of the ICD-10 diagnosis of PCOS (E282) or anovulatory infertility (N970) in the Swedish Patient Registry. The PCOS diagnosis during the study period in Sweden was made mainly according to the 2003 Rotterdam criteria for PCOS, but according to National Guidelines, stricter criteria were in use for adolescents. The revised Rotterdam criteria demanded two out of the following three features, that include the following: (i) oligo-/anovulation, (ii) clinical and/or biochemical hyperandrogenism and (iii) polycystic ovarian morphology on ultrasound, together with exclusion of other etiologies (The Rotterdam ESHRE/ASRM-SPCWG, 2004). In adolescence, PCOS diagnosis was made according to the NIH PCOS criteria which required the presence of both clinical and/or biochemical hyperandrogenism and chronic anovulation, after other etiologies were excluded (such as androgen-secreting tumors, Cushing's syndrome and congenital adrenal hyperplasia) (Zawadski and Dunaif, 1992; Rosenfield, 2020). Women diagnosed with anovulatory infertility were also included in our study population based on the fact that 90% of them have PCOS, according to the Rotterdam criteria (Broekmans *et al.*, 2006; Teede *et al.*, 2010).

Outcome

The outcome measure was first childbirth after spontaneous conception which was considered as a proxy for restoration of normal fertility. Information on fertility surgery, ovulation induction, assisted reproduction, IVF and other infertility treatments were recorded at the first antenatal visit by use of check boxes and were retrieved from the Medical Birth Registry. Information on first childbirth was collected and classified as spontaneous conception if no form of assisted reproduction had been recorded. Time to childbirth is estimated in years from the time a participant turned 18 until the year of first childbirth or the end of the follow-up period.

Study population

The initial population included women born between 1971 and 1997 according to the presence or absence of PCOS or anovulatory infertility diagnosis (Persson *et al.*, 2019). The control group comprised five control individuals per each woman with PCOS, matched by year of birth and residential area, randomly chosen from the Total Population Registry. All women of the study or control group with hyperprolactinemia (E221), congenital adrenal hyperplasia (E25), premature ovarian insufficiency (E283) or Turner syndrome (Q96) were excluded from the population. Lastly, women with one or more prior births before the first recorded birth in the Medical Birth Registry were also excluded.

Due to limitations in data availability from the Swedish Prescribed Drug Registry (i.e. registry founded in 2005), we further restricted our population to women with access to data regarding the dispensing of medications during adolescence, i.e. born between 1987 and 1996. Since the aim of the study was to explore the incidence of the outcome occurring after the intervention (definition follows), all participants with births registered before 18 years of age were excluded from the population (Morgan, 2019). In the end, 15 106 women with PCOS and 73 786 control women were eligible for inclusion in the study.

Intervention

The intervention of interest regarded the *early use* of commonly prescribed anti-androgenic treatment (Bremer, 2010; Ibáñez *et al.*, 2017; Witchel *et al.*, 2019) comprising certain combined oral contraceptives (COCs) and/or other anti-androgens (detailed description follows). Anti-androgenic medications were promoted in adolescents

and adults with PCOS and hirsutism during the study period, based on Swedish National Guidelines (Swedish Medical Products Agency (Läkemedelsverket), 2014). Notably, this is no longer the case (Teede et al., 2018). Information on anti-androgenic treatment was retrieved from the Swedish Prescribed Drug Registry with the use of ATC codes. The medications of interest included the following: (i) selected COCs advocated against hyperandrogenism by the Swedish contraceptive policy guidelines (referred to as anti-androgenic COCs in the study or aaCOCs) (Swedish Medical Products Agency (Läkemedelsverket), 2014), such as ethinylestradiol (EE) and dienogest (G03AA16), EE and drosperinone (G03AA12), EE and desogestrel (G03AA09); and/or (ii) prescription of other anti-androgenic medications such as spironolactone (C03DA01), finasteride and dutasteride (G04CB), finasteride and eflornithine (D11AX), flutamide and bikalutamide (L02BB), EE and cyproterone acetate (G03HB01). Women classified as treated received at least two filled prescriptions of any of the medications listed above during or after adolescence.

Due to the 'registry-based' study design, data on the clinical or biochemical androgen status of PCOS women (Lizneva et al., 2016) are lacking. Instead, the prescription of anti-androgenic medications was used as a proxy for evidence of hyperandrogenism. Women with PCOS who were not treated with any anti-androgenic medication were regarded as normo-androgenic, while those treated were regarded as hyperandrogenic. We further classified hyperandrogenic women as being mildly hyperandrogenic if they received aaCOC monotherapy, or severely hyperandrogenic if they received other anti-androgens with or without aaCOCs. Early anti-androgenic treatment was defined as during adolescence ($\leq$18 years of age) (referred to as early users) or after adolescence (>18 years) (referred to as late users). A flow diagram of the study design is shown in Supplementary Fig. S1.

Covariates

Data on obesity were retrieved from the Swedish Patient Registry and concerned the presence of the ICD-10 diagnosis on obesity (E66) (i.e. BMI $\geq$ 30 kg/m^2). Data on year and country of birth, as well as maternal education in 2017, were retrieved from the Total Population Registry and the Education Registry, respectively. Maternal country of birth was categorized as Nordic (including Sweden, Finland, Denmark, Norway and Iceland), European (excluding Nordic countries), Middle Eastern, South Asian (India, Bangladesh and Pakistan), African and remaining countries. Maternal education was categorized as <12 or $\geq$12 years.

Statistical analysis

The statistical analyses were performed using the Statistical Package for the Social Sciences (SPSS) (IBM Corp., Armonk, NY, USA, Version 26). The probability of first childbirth after spontaneous conception was analyzed with the use of Kaplan–Meier hazard curve. Using the Cox proportional hazards regression test with time-dependent covariates, we ensured that the assumption of proportional hazards was fulfilled. The average time to childbirth, calculated only among women with the end-point event, was presented with both mean and median calculated values, and statistical comparisons were made with the use of the log-rank test. When a participant reached the end of the observation period, censoring was applied. We used the landmark analysis (Morgan, 2019), restricting the results to those women still at risk at the landmark time (i.e. 18 years of age) and ignoring all those with the event prior to the landmark. In order to avoid bias, the landmark was chosen based on clinical relevance, prior to the data analysis. We estimated the FR and 95% confidence interval for the time to first childbirth after spontaneous conception, using Cox proportional hazards regression models. FR below 1.0 (< 1.0) denotes reduced fecundity for the group of interest compared to the reference group. The Cox regression analyses were adjusted for obesity, birth year, country of birth and education level. The Cox regression analyses concern comparisons between the study and control group, as well different PCOS subcategories (normo-androgenic, mildly hyperandrogenic or severely hyperandrogenic women, and early users or late users). A subgroup analysis among hyperandrogenic PCOS women in relation to the timing of treatment initiation (early versus late) was performed after stratification on the severity of hyperandrogenism. Lastly, sensitivity analyses were performed restricting the study population to (i) women with PCOS diagnosis only (ICD-10 code E282), (ii) women with PCOS of Nordic origin only, as well as (iii) women with PCOS on COCs.

Results

Sociodemographic characteristics

The background characteristics of the total population are presented in Table I. Greater proportions of women originating from Europe, the Middle East and South Asia, as well as women with lower education level (below 12 years) were observed in the PCOS study group. Furthermore, the rate of obesity was higher in women with PCOS compared to the non-PCOS controls (13.3% versus 3.4%, $P < 0.001$). Women with PCOS also had a significantly lower incidence of childbirth after spontaneous conception (14.2% versus 18.9%, $P < 0.001$) compared to the non-PCOS controls. More than half of the women with PCOS (n = 7 949, 52.6%) had been dispensed anti-androgenic medications. The most commonly prescribed anti-androgenic medications in women with PCOS were aaCOCs, either as monotherapy (n = 5 456, 36.1%) or in combination with plain anti-androgens (n = 1 533, 10.1%). Plain anti-androgen monotherapy was less common (n = 960, 6.4%). The medications were most commonly prescribed after adolescence (late users) (71.4%) compared to during adolescence (early users) (28.6%). A higher proportion of normo-androgenic women with PCOS gave birth following a spontaneous conception compared to hyperandrogenic women with PCOS (17.1% versus 11.6%, respectively) (Table II).

Non-PCOS controls have a greater probability of spontaneous childbirth after spontaneous conception than normo-androgenic and hyperandrogenic women with PCOS

In comparison with non-PCOS controls, the probability of childbirth after spontaneous conception was 11% lower among normo-androgenic women with PCOS, and 40% lower among hyperandrogenic women

Table I Background characteristics of the total population (N = 88 892) including women with and without polycystic ovary syndrome (PCOS).

	Women with PCOS (n = 15 106)	Non-PCOS controls (n = 73 786)	P-value
Obesity			<0.001
No	13 102 (86.7)	71 302 (96.6)	
Yes	2004 (13.3)	2484 (3.4)	
Education			<0.001
<12 years	7701 (51.7)	35 812 (49.6)	
≥12 years	7181 (48.3)	36 368 (50.4)	
Missing data (1830, 2.1)			
Country of Birth			<0.001
Nordic countries	11 500 (76.1)	61 318 (83.1)	
Europe	1087 (7.2)	4526 (6.1)	
Middle East	1425 (9.4)	2872 (3.9)	
South Asia	260 (1.7)	540 (0.7)	
Africa	263 (1.7)	1706 (2.4)	
Remaining countries	571 (3.8)	2824 (3.8)	
Birth year			NS
1987	2361 (15.6)	11 401 (15.5)	
1988	2210 (14.6)	10 738 (14.6)	
1989	2084 (13.8)	10 153 (13.8)	
1990	1951 (12.9)	9523 (12.9)	
1991	1704 (11.3)	8304 (11.3)	
1992	1408 (9.3)	6936 (9.4)	
1993	1141 (7.6)	5642 (7.6)	
1994	940 (6.2)	4648 (6.3)	
1995	760 (5.0)	3745 (5.1)	
1996	547 (3.6)	2696 (3.7)	
First spontaneous childbirth			<0.001
No	12 961 (85.8)	59 856 (81.1)	
Yes	2145 (14.2)	13 930 (18.9)	

Data are presented as n (%).

Table II Treatment characteristics of the study group of polycystic ovary syndrome (PCOS) women.

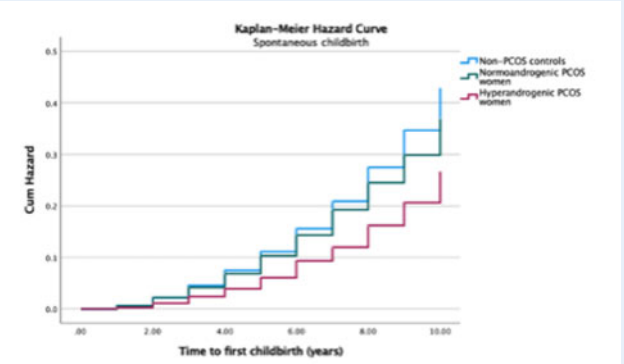
	Women with PCOS (n = 15 106)
Hyperandrogenic status	
Normo-androgenic PCOS women	7157 (47.4%)
Hyperandrogenic PCOS women	7949 (52.6%)
Combinations of aa medications used	
No anti-androgenic medications	7157 (47.4%)
aaCOCs only	5456 (36.1%)
aaCOCs and plain anti-androgens combined	1533 (10.1%)
Plain anti-androgens only	960 (6.4%)
Hyperandrogenic status	
Normo-androgenic PCOS women	7157 (47.4%)
Mildly hyperandrogenic PCOS women	5456 (36.1%)
Severely hyperandrogenic PCOS women	2493 (16.5%)
Timing of any anti-androgenic medications**	
Early users	2276 (15.1%)
Late users	5673 (37.6%)
Non users	7157 (47.4%)
aaCOCs	
Early users	1939 (12.8%)
Late users	5050 (33.4%)
Non users	8117 (53.7%)
Plain anti-androgens	
Early users	514 (3.4%)
Late users	1979 (13.1%)
Non users	12 613 (83.5%)

**Early and late users comprise PCOS women with anti-androgenic treatment initiated up to or above 18 years of age, respectively.
aaCOCs, anti-androgenic combined oral contraceptives.
Data are presented as n (%).

with PCOS (unadjusted FR 0.89 (95% CI 0.84–0.95) and unadjusted FR 0.60 (95% CI 0.56–0.64), respectively). The estimates remained unchanged after adjustment for obesity, year of birth, country of birth and education level (Fig. 1). The calculated mean and median time to childbirth after spontaneous conception were shortest in normo-androgenic women with PCOS (mean 5.51 years, SD 2.35/median 6.0 years, IQR 3.0) compared to non-PCOS controls (mean 5.60 years, SD 2.37/median 6.0 years, IQR 3.0) and hyperandrogenic women with PCOS (mean 5.84 years, SD 2.34/median 6.0 years, IQR 4.0) ($P = 0.005$).

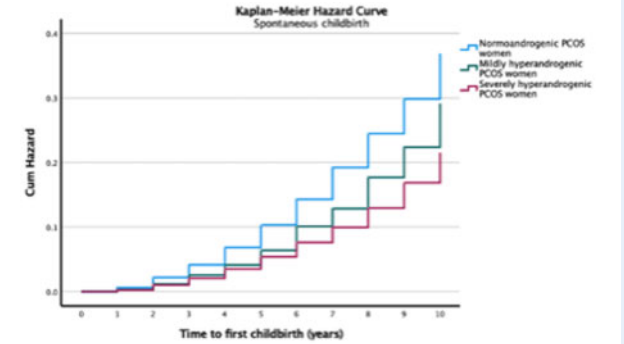
Severely hyperandrogenic women with PCOS have a lower probability of childbirth after spontaneous conception compared to mildly hyperandrogenic women with PCOS

Compared to normo-androgenic women with PCOS, we observed that the FR was lowest among severely hyperandrogenic women with PCOS (unadjusted FR 0.58 (95% CI 0.51–0.66)), followed by mildly hyperandrogenic women with PCOS (unadjusted FR 0.72 (95% CI 0.65–0.79)) (Fig. 2). The above estimates did not change after adjustment for obesity, year of birth, country of birth and education level. Mean and median time to first childbirth after spontaneous conception was significantly longer in severely hyperandrogenic women with PCOS (mean 5.87 years, SD 2.36/median 6.0 years, IQR 4.0) compared to mildly hyperandrogenic women with PCOS (mean 5.83 years, SD 2.34/median 6.0 years, IQR 4.0) and normo-androgenic women



	Crude model	Adjusted model
	FR (95% CI)	FR (95% CI)
Non-PCOS controls	Ref	Ref
Normo-androgenic PCOS women	0.89 (0.84-0.95)	0.68 (0.64 - 0.72)
Hyperandrogenic PCOS women	0.60 (0.56-0.64)	0.53 (0.50 - 0.57)

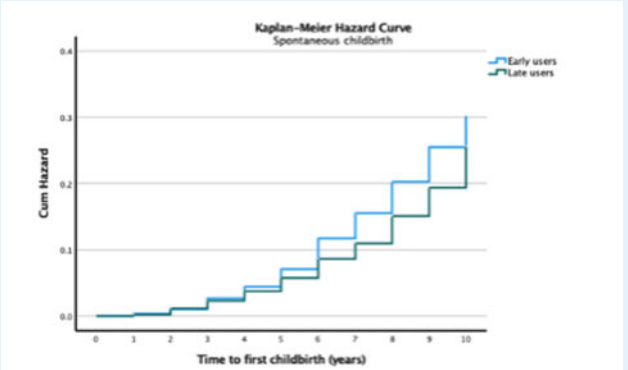
Figure 1. Probability of first childbirth by spontaneous conception in the entire population. Cum, cumulative; PCOS, polycystic ovary syndrome.



	Crude model	Adjusted model
	FR (95% CI)	FR (95% CI)
Normo-androgenic PCOS controls	Ref	Ref
Mildly hyperandrogenic PCOS women	0.72 (0.66 – 0.79)	0.84 (0.77 – 0.93)
Severely hyperandrogenic PCOS women	0.55 (0.48 – 0.63)	0.60 (0.52 – 0.69)

Figure 2. Probability of first childbirth by spontaneous conception among polycystic ovary syndrome (PCOS) women in relation to the anti-androgenic potential of medication used.

with PCOS (mean 5.51 years, SD 2.35/median 6.0 years, IQR 3.0) ($P = 0.007$).



	Crude model	Adjusted model
	FR (95% CI)	FR (95% CI)
Early users	Ref	Ref
Late users	0.77 (0.67 – 0.90)	0.79 (0.68 – 0.92)

Figure 3. Probability of first childbirth by spontaneous conception among hyperandrogenic polycystic ovary syndrome (PCOS) women in relation to the timing of medication used.

Early anti-androgenic treatment in women with PCOS is associated with a higher probability of childbirth after spontaneous conception compared to late anti-androgenic treatment

In comparison with women with PCOS who started anti-androgenic treatment in adolescence, late users exhibited a lower FR (unadjusted FR 0.78 (95% CI 0.67–0.90)) with similar estimates even after adjustment (Fig. 3). Mean and median time to first childbirth after spontaneous conception was shorter in hyperandrogenic women with PCOS who started anti-androgenic treatment early (mean 5.40 years, SD 2.15/median 6.0 years, IQR 3.0) compared to women who started late (mean 6.00 years, SD 2.39/median 6.0 years, IQR 4.0) ($P = 0.001$). Similar results regarding the effect of early or late treatment initiation could be seen even after stratifying the PCOS population on the severity of hyperandrogenism. Early treatment was especially effective among mildly hyperandrogenic PCOS women (unadjusted FR 0.72 (95% CI 0.61–0.86)), and to a lesser degree in severely hyperandrogenic women (unadjusted FR 0.89 (95% CI 0.67–1.19)).

Numbers at risk over time in each model are shown in Supplementary Table SI.

Sensitivity analyses regarding all three models/scenarios presented above were performed among women with PCOS diagnosis (ICD-10 code E282) (i.e. without taking into account whether women were also diagnosed with anovulatory infertility) as well as PCOS women of Nordic origin or women on aaCOCs, and none of them altered the estimates presented previously.

Discussion

Main findings

As expected, non-PCOS controls had a higher probability of childbirth after spontaneous conception than normo-androgenic and

hyperandrogenic PCOS women. Among women with PCOS, severe hyperandrogenism (characterized by the higher anti-androgenic potential of treatment) was associated with lower FR, while the initiation of anti-androgenic treatment at an earlier age was associated with ameliorated severity of PCOS-related subfertility.

The need for anti-androgenic treatment likely corresponds to individuals with more severe hyperandrogenic clinical manifestations. It is therefore not surprising that PCOS-diagnosed women who received anti-androgen treatment (at some point in life) required more time to first childbirth after a spontaneous conception and had a lower FR in comparison with non-PCOS controls and normo-androgenic women with PCOS. Initiation of anti-androgenic treatment at an earlier stage (i.e. during adolescence rather than later in life) suggests early awareness of PCOS, which, in turn, may affect reproductive choices and family planning. However, we consider anti-androgenic treatment at an earlier age as an unlikely marker of a deliberate attempt at achieving pregnancy since anti-androgenic drugs either are contraceptives or are teratogenic (Eibs *et al.*, 1982; Kim and Del Rosso, 2012). One could therefore wonder whether the higher FR among early users observed in our study could be attributed to a long-lasting pharmacological effect of the anti-androgens, extending even after the period of use.

Comparison to other studies

While there is a paucity of clinical data on the long-term impact of anti-androgen treatment of women with PCOS, the present findings are in agreement with what is reported. Long-term androgen receptor blockade is associated with improved testosterone levels and ovulatory function in adult women with PCOS (De Leo *et al.*, 1998; Paradisi *et al.*, 2013). Additionally, long-term anti-androgen therapy is reported to restore normal steroid hormone feedback to the reproductive axis (Eagleson *et al.*, 2000). However, there is a lack of prospective clinical data reporting the FR of PCOS women taking anti-androgens.

Anti-androgen administration in animals that mimic the PCOS phenotype following prenatal exposure to androgen excess can restore normal estrous cyclicity (Sullivan and Moenter, 2004; Silva *et al.*, 2018). Endogenous hyperandrogenism increases in this model at 40 days, shortly after pubertal onset (Silva *et al.*, 2018). Flutamide treatment from 40 to 60 days rescues estrous cyclicity, improves ovarian morphology and also reverses aberrant GABA wiring in the brain associated with impaired steroid hormone feedback of PCOS (Moore *et al.*, 2015; Silva *et al.*, 2018). It remains to be determined whether these improvements are the same if treatment is delayed beyond this 'adolescent' period.

Pathophysiology

These data suggest that by 'hindering or impairing' the 'second hit' of androgen action that develops at puberty, particularly through early pharmacological intervention, PCOS-related subfertility may be attenuated. The critical site of androgen action in the development and maintenance of clinical PCOS features remains unclear, but several lines of evidence in preclinical animal models point toward the importance of androgen signaling in the female brain (Coutinho and Kauffman, 2019; Ruddenklau and Campbell, 2019). Given the multifactorial nature of PCOS pathogenesis, early interventions of 'second hits' are likely to be most effective when guided by known predisposing contributors, or 'first hits'(Ibáñez *et al.*, 2017).

One possibility in women with PCOS is that the 'first hit' (i.e. genetic predisposition, prenatal androgen excess) shapes hypothalamic circuitry in a particular way that establishes the emergence of PCOS features, including aberrant gonadotrophin regulation and hyperandrogenism, and that a 'second hit' is then required to maintain those pathophysiological changes. Of interest, changes in GABA brain wiring and activity associated with prenatal androgen excess in rodent models develop prior to hyperandrogenism and PCOS-like features (Berg *et al.*, 2018; Silva *et al.*, 2018). However, these 'programmed' changes in neuroendocrine circuits can be rewired with androgen blockade (Silva *et al.*, 2018).

Strengths and limitations

The main strength of the study is its large sample size which enabled us to perform analyses of rare outcomes, as well as stratify according to different variables. Furthermore, information was retrieved from highly valid registries and evolved over a period of 10 years.

Unfortunately, we lacked information on smoking or possible comorbidities of the study participants. In addition, since the BMI of study participants is not available unless they give birth, we employed the presence of obesity diagnosis instead. While the dependence on ICD diagnosis will likely under-report obesity rates among this cohort, this enabled us to account for obesity in the adjusted Cox regression models to overcome the limitation of not having access to BMI data. Furthermore, it would have been desirable to have accurate and direct information on the androgenic biochemical status of the PCOS participants; we employed instead use of anti-androgenic medications as a proxy of hyperandrogenism. Moreover, we do not know the duration of anti-androgenic treatment utilized or the number of study participants that actively attempted a pregnancy during the study period. In addition, we did not censor for death or immigration, and all participants were followed up until the ending date. Lastly, it has been demonstrated that foreign-born families residing in Sweden utilize healthcare services with increased frequency compared to natives, increasing the chance that a PCOS diagnosis is posed and thereby treatment initiated at an earlier stage (Swedish Public Health Agency (Folkhälsomyndigheten), 2019). Furthermore, women with a foreign background attempt pregnancy at an earlier age compared to Swedish-born women (National Board of Health and Welfare (Socialstyrelsen), 2020). Both observations could indicate that the differences noted may be due to social reasons and not only biological reasons. However, our results remained unaltered, even after the sensitivity analysis performed restricting our population to women of Nordic origin, strengthening the biological component of the intervention.

Lastly, women with PCOS, especially those with more severe clinical features of the syndrome such as the obese and/or hyperandrogenic women, often report distorted self-perceived body image which in turn affects their sexuality and social well-being (Alur-Gupta *et al.*, 2019; Kogure *et al.*, 2019). Whether body dissatisfaction results in a deliberate delay in childbearing is unknown and difficult to tease apart from the PCOS-associated impairments in the reproductive axis. The possibility of bias cannot therefore be entirely ruled out, but it is deemed nevertheless to be limited since the effect of hyperandrogenism on childbirth is not attenuated after adjustment for obesity.

Conclusions

Our study findings suggest that early initiation of anti-androgenic treatment during adolescence could promote the prevention of fertility-related morbidity among PCOS women. Although reproductive endocrinologists should exercise caution in assigning PCOS diagnosis prematurely, they should however not stall early treatment initiation when indicated. These findings support the need for future interventional randomized prospective studies investigating critical windows of anti-androgen treatment.

Supplementary data

Supplementary data are available at *Human Reproduction* online.

Data availability

The data sets generated and analyzed during the current study are not publicly available and cannot be uploaded at any website due to the risk of compromising the individual privacy of participants.

Authors' roles

All authors have equally contributed to the design of the study, data collection, interpretation of the results, drafting of the manuscript, as well as critical revision of the final version submitted.

Funding

This study was funded by the Health Research Council of New Zealand (18-671), the Swedish Society of Medicine and the Uppsala University Hospital.

Conflict of interest

Evangelia Elenis has over the past year received lecture fee from Gedeon Richter outside the submitted work. Inger Sundström Poromaa has over the past three years received compensation as a consultant and lecturer for Bayer Schering Pharma, MSD, Gedeon Richter, Peptonics and Lundbeck A/S. The other authors have no conflict of interests to declare.

References

Alur-Gupta S, Chemerinski A, Liu C, Lipson J, Allison K, Sammel MD, Dokras A. Body-image distress is increased in women with polycystic ovary syndrome and mediates depression and anxiety. *Fertil Steril* 2019;**112**:930–938.e1.

Berg T, Silveira MA, Moenter SM. Prepubertal development of GABAergic transmission to gonadotropin-releasing hormone (GnRH) neurons and postsynaptic response are altered by prenatal androgenization. *J Neurosci* 2018;**38**:2283–2293.

Bremer AA. Polycystic ovary syndrome in the pediatric population. *Metab Syndr Relat Disord* 2010;**8**:375–394.

Broekmans FJ, Knauff EA, Valkenburg O, Laven JS, Eijkemans MJ, Fauser BC. PCOS according to the Rotterdam consensus criteria: change in prevalence among WHO-II anovulation and association with metabolic factors. *BJOG* 2006;**113**:1210–1217.

Coutinho EA, Kauffman AS. The role of the brain in the pathogenesis and physiology of polycystic ovary syndrome (PCOS). *Med Sci (Basel)* 2019;**7**:84.

De Leo V, Lanzetta D, D'Antona D, la Marca A, Morgante G. Hormonal effects of flutamide in young women with polycystic ovary syndrome. *J Clin Endocrinol Metab* 1998;**83**:99–102.

de Zegher F, Lopez-Bermejo A, Ibanez L. Central obesity, faster maturation, and 'PCOS' in girls. *Trends Endocrinol Metab* 2018;**29**:815–818.

Driscoll DA. Polycystic ovary syndrome in adolescence. *Semin Reprod Med* 2003;**21**:301–307.

Eagleson CA, Gingrich MB, Pastor CL, Arora TK, Burt CM, Evans WS, Marshall JC. Polycystic ovarian syndrome: evidence that flutamide restores sensitivity of the gonadotropin-releasing hormone pulse generator to inhibition by estradiol and progesterone. *J Clin Endocrinol Metab* 2000;**85**:4047–4052.

Eibs HG, Spielmann H, Jacob-Muller U, Klose J. Teratogenic effects of cyproterone acetate and medroxyprogesterone treatment during the pre- and postimplantation period of mouse embryos. II. Cyproterone acetate and medroxyprogesterone acetate treatment before implantation in vivo and in vitro. *Teratology* 1982;**25**:291–299.

Ibáñez L, Oberfield SE, Witchel S, Auchus RJ, Chang RJ, Codner E, Dabadghao P, Darendeliler F, Elbarbary NS, Gambineri A *et al*. An international consortium update: pathophysiology, diagnosis, and treatment of polycystic ovarian syndrome in adolescence. *Horm Res Paediatr* 2017;**88**:371–395.

Kim GK, Del Rosso JQ. Oral spironolactone in post-teenage female patients with acne vulgaris: practical considerations for the clinician based on current data and clinical experience. *J Clin Aesthet Dermatol* 2012;**5**:37–50.

Kogure GS, Ribeiro VB, Lopes IP, Furtado CLM, Kodato S, Silva de Sa MF, Ferriani RA, Lara L, Maria Dos Reis R. Body image and its relationships with sexual functioning, anxiety, and depression in women with polycystic ovary syndrome. *J Affect Disord* 2019;**253**:385–393.

Koivunen R, Pouta A, Franks S, Martikainen H, Sovio U, Hartikainen AL, McCarthy MI, Ruokonen A, Bloigu A, Jarvelin MR. Fecundability and spontaneous abortions in women with self-reported oligo-amenorrhea and/or hirsutism: Northern Finland Birth Cohort 1966 Study. *Hum Reprod* 2008;**23**:2134–2139.

Lizneva D, Suturina L, Walker W, Brakta S, Gavrilova-Jordan L, Azziz R. Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertil Steril* 2016;**106**:6–15.

Ludvigsson JF, Andersson E, Ekbom A, Feychting M, Kim JL, Reuterwall C, Heurgren M, Olausson PO. External review and validation of the Swedish national inpatient register. *BMC Public Health* 2011;**11**:450.

Ludvigsson JF, Otterblad-Olausson P, Pettersson BU, Ekbom A. The Swedish personal identity number: possibilities and pitfalls in healthcare and medical research. *Eur J Epidemiol* 2009;**24**:659–667.

March WA, Moore VM, Willson KJ, Phillips DI, Norman RJ, Davies MJ. The prevalence of polycystic ovary syndrome in a community sample assessed under contrasting diagnostic criteria. *Hum Reprod* 2010;**25**:544–551.

Moore AM, Prescott M, Campbell RE. Estradiol negative and positive feedback in a prenatal androgen-induced mouse model of polycystic ovarian syndrome. *Endocrinology* 2013;**154**:796–806.

Moore AM, Prescott M, Marshall CJ, Yip SH, Campbell RE. Enhancement of a robust arcuate GABAergic input to gonadotropin-releasing hormone neurons in a model of polycystic ovarian syndrome. *Proc Natl Acad Sci USA* 2015;**112**:596–601.

Morgan CJ. Landmark analysis: a primer. *J Nucl Cardiol* 2019;**26**:391–393.

National Board of Health and Welfare [Socialstyrelsen]. Statistics on pregnancies, childbirths and infants born 2018. [Statistik om graviditeter, förlossningar och nyfödda barn 2018], 2020.

Paradisi R, Fabbri R, Battaglia C, Venturoli S. Ovulatory effects of flutamide in the polycystic ovary syndrome. *Gynecol Endocrinol* 2013;**29**:391–395.

Persson S, Elenis E, Turkmen S, Kramer MS, Yong EL, Sundstrom-Poromaa I. Fecundity among women with polycystic ovary syndrome (PCOS)-a population-based study. *Hum Reprod* 2019;**34**:2052–2060.

Rosenfield RL. Current concepts of polycystic ovary syndrome pathogenesis. *Curr Opin Pediatr* 2020;**32**:698–706.

Rosenfield RL, Ehrmann DA. The pathogenesis of polycystic ovary syndrome (PCOS): the hypothesis of PCOS as functional ovarian hyperandrogenism revisited. *Endocr Rev* 2016;**37**:467–520.

Ruddenklau A, Campbell RE. Neuroendocrine impairments of polycystic ovary syndrome. *Endocrinology* 2019;**160**:2230–2242.

Ryan GE, Malik S, Mellon PL. Antiandrogen treatment ameliorates reproductive and metabolic phenotypes in the letrozole-induced mouse model of PCOS. *Endocrinology* 2018;**159**:1734–1747.

Silva MS, Prescott M, Campbell RE. Ontogeny and reversal of brain circuit abnormalities in a preclinical model of PCOS. *JCI Insight* 2018;**3**:e99405.

Sullivan SD, Moenter SM. Prenatal androgens alter GABAergic drive to gonadotropin-releasing hormone neurons: implications for a common fertility disorder. *Proc Natl Acad Sci USA* 2004;**101**:7129–7134.

Swedish Medical Products Agency [Läkemedelsverket]. Constraception – treatment recommendations. [Antikonception – behandlingsrekommendation]. 2014;**25**:14–28.

Swedish Public Health Agency [Folkhälsomyndigheten]. Health, socioeconomic, and lifestyle factors among foreign born individuals in Sweden. [Hälsa hos personer som är utrikes födda – skillnader i hälsa utifrån födelseland. 2019.

Teede H, Deeks A, Moran L. Polycystic ovary syndrome: a complex condition with psychological, reproductive and metabolic manifestations that impacts on health across the lifespan. *BMC Med* 2010;**8**:41.

Teede HJ, Misso ML, Costello MF, Dokras A, Laven J, Moran L, Piltonen T, Norman RJ, Andersen M, Azziz R, International PCOS Network *et al*. Recommendations from the international evidence-based guideline for the assessment and management of polycystic ovary syndrome. *Hum Reprod* 2018;**33**:1602–1618.

The Rotterdam ESHRE/ASRM-SPCWG. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil Steril* 2004;**81**:19–25.

West S, Vahasarja M, Bloigu A, Pouta A, Franks S, Hartikainen AL, Jarvelin MR, Corbett S, Vaarasmaki M, Morin-Papunen L. The impact of self-reported oligo-amenorrhea and hirsutism on fertility and lifetime reproductive success: results from the Northern Finland Birth Cohort 1966. *Hum Reprod* 2014;**29**:628–633.

Witchel SF, Oberfield SE, Pena AS. Polycystic ovary syndrome: pathophysiology, presentation, and treatment with emphasis on adolescent girls. *J Endocr Soc* 2019;**3**:1545–1573.

Wolf WM, Wattick RA, Kinkade ON, Olfert MD. Geographical prevalence of polycystic ovary syndrome as determined by region and race/ethnicity. *Int J Environ Res Public Health* 2018;**15**:2589.

Zawadski J, Dunaif A. Diagnostic criteria for polycystic ovary syndrome; towards a rational approach. In: A Dunaif, J Givens, F Haseltine (eds). *Polycystic Ovary Syndrome*. Boston, MA: Black-Well Scientific, 1992, 377–384.

THE JOURNAL OF
SEXUAL MEDICINE

ORIGINAL RESEARCH & REVIEWS

TRANSGENDER HEALTH

Mental Healthcare Utilization of Transgender Youth Before and After Affirming Treatment

Elizabeth Hisle-Gorman, MSW, PhD,[1] Natasha A. Schvey, PhD,[2] Terry A. Adirim, MD, MPH,[3] Anna K. Rayne, MD,[4] Apryl Susi, MS,[5] Timothy A. Roberts, MD, MPH,[6] and David A. Klein, MD, MPH[7]

ABSTRACT

Objective: Transgender and gender-diverse (TGD) adolescents experience increased mental health risk compared to cisgender peers. Limited research suggests improved outcomes following gender-affirmation. This study examined mental healthcare and psychotropic medication utilization among TGD youth compared to their siblings without gender-related diagnoses and explored utilization patterns following gender-affirming care.

Method: This retrospective cohort study used military healthcare data from 2010−2018 to identify mental healthcare diagnoses and visits, and psychotropic medication prescriptions among TGD youth who received care for gender dysphoria before age 18, and their siblings. Logistic and Poisson regression analyses compared mental health diagnosis, visits, and psychotropic prescriptions of TGD youth to their siblings, and compared healthcare utilization pre- and post-initiation of gender-affirming pharmaceuticals among TGD adolescents.

Results: 3,754 TGD adolescents and 6,603 cisgender siblings were included. TGD adolescents were more likely to have a mental health diagnosis (OR 5.45, 95% CI [4.77−6.24]), use more mental healthcare services (IRR 2.22; 95% CI [2.00−2.46]), and be prescribed more psychotropic medications (IRR = 2.57; 95% CI [2.36−2.80]) compared to siblings. The most pronounced increases in mental healthcare were for adjustment, anxiety, mood, personality, psychotic disorders, and suicidal ideation/attempted suicide. The most pronounced increased in psychotropic medication were in SNRIs, sleep medications, anti-psychotics and lithium. Among 963 TGD youth (M_{age}: 18.2) using gender-affirming pharmaceuticals, mental healthcare did not significantly change (IRR = 1.09, 95% CI [0.95−1.25]) and psychotropic medications increased (IRR = 1.67, 95% CI [1.46−1.91]) following gender-affirming pharmaceutical initiation; older age was associated with decreased care and prescriptions.

Conclusion: Results support clinical mental health screening recommendations for TGD youth. Further research is needed to elucidate the longer-term impact of medical affirmation on mental health, including family and social factors associated with the persistence and discontinuation of mental healthcare needs among TGD youth.
Hisle-Gorman E, Schvey NA, Adirim TA, et al. Mental Healthcare Utilization of Transgender Youth Before and After Affirming Treatment. J Sex Med 2021;18:1444−1454.

Key Words: Transgender; Gender-Diverse; Mental Health; Adolescent; Youth

Received December 30, 2020. Accepted May 14, 2021.

[1]Department of Pediatrics, Uniformed Services University, Bethesda, MD; Department of Preventive Medicine and Biostatistics, Uniformed Services University, Bethesda, MD, USA;

[2]Department of Medical and Clinical Psychology, Uniformed Services University, Bethesda, MD, USA;

[3]Department of Preventive Medicine and Biostatistics, Uniformed Services University, Bethesda, MD; Department of Defense, Washington, DC, USA;

[4]Department of Family Medicine, Fort Belvoir Community Hospital, Fort Belvoir, VA, USA;

[5]Department of Pediatrics, Uniformed Services University, Bethesda, MD; Department of Preventive Medicine and Biostatistics, Uniformed Services University, Bethesda, MD, USA;

[6]Division of Adolescent Medicine, Children's Mercy Kansas City; Department of Pediatrics, University of Missouri-Kansas City School of Medicine, Kansas City, MO, USA;

[7]Department of Pediatrics, Uniformed Services University, Bethesda, MD; Department of Family Medicine, Uniformed Services University, Bethesda, MD, USA

https://doi.org/10.1016/j.jsxm.2021.05.014

BACKGROUND

Transgender and gender-diverse (TGD) youth include those whose gender identity, expression, or behavior differs from that typically associated with their sex assigned at birth.[1] An estimated 0.7−2.7 percent of adolescents identify as TGD,[2−5] and TGD individuals are increasingly presenting for associated healthcare.[1,6−10]

While TGD individuals remain under-represented in medical research,[11−13] a growing body of literature suggests significant health disparities and poorer mental and physical health among TGD individuals as compared to cisgender peers.[2,3,14−21] In adult populations, large data studies of TGD Medicare recipients and Veterans indicate TGD adults use more mental healthcare, and experience increased disability, chronic conditions, substance abuse disorders, chronic pain, suicide, suicide related events, and disabling mental illness as compared to cisgender controls.[14−18] Health outcomes appear related to environment, with Veterans living in more accepting communities having fewer substance use and mental health comorbidities.[22,23]

In studies of children and adolescents, school and internet based surveys in the United States and other parts of the world, found self-identified TGD youth were less likely to report having a caring parent, and more likely to report depression, suicide attempts, suicidal risk, violence victimization, self-harm, substance use, unsafe sex, psychological distress, and bullying as compared to cisgender peers − outcomes likely related to stigma, family rejection, and victimization.[2,3,19,20,24−29] Parents of 105 adolescents with gender dysphoria reported 32% had a concurrent psychiatric disorder, including anxiety, mood, and disruptive disorders, and that multiple diagnoses were increased in those with transfeminine identities.[30] Larger studies using records from a community-based clinic and a 2-state integrated health system compared TGD youth to cisgender controls and found that the odds of multiple mental health diagnoses were increased 2 to several fold in TGD youth.[21,31]

However, some research indicates that mental health conditions in TGD youth and adults were not elevated, or were ameliorated in those with some level of medical or social affirmation (ie, those supported to live openly in their asserted gender identity).[32−43] Limited data, primarily using small samples, and self-report measures, indicate that mental health concerns and suicidality decreased, and well-being increased, following medical or social affirmation.[32−41,61] Small studies of youth who have completed social affirmation report improvement on psychological functioning and well-being and decreased gender dysphoria, but generally results rely on parent or child-report of symptoms.[35,36] Among adults, a meta-analysis of 1,833 TGD adults indicated self-reported improvement on gender dysphoria, psychological symptoms, quality of life, and sexual function following pharmaceutical affirmation.[44] One study in adults found that length of hormone treatment was not associated with changes in healthcare utilization for mood or anxiety disorders, but time from surgery was associated with decreases in care.[41,61]

Most TGD pediatric mental health research is limited by use of self- or parent-report, small sample size, limited geographic area, and lack of a non-TGD control group. While 2 studies have explored mental health diagnoses in larger samples,[21,31] neither examined patterns in mental healthcare utilization nor psychotropic medication prescriptions, which are important indicators of the severity of mental health conditions. Research specifically exploring effects of gender-affirming care on mental health have similar limitations, with few including adolescents or young adults.[32−41,61]

Given research indicating that TGD youth may be at increased risk for mental health conditions adequately powered studies are needed to better elucidate the mental healthcare needs of TGD youth, compared to matched controls, and to identify the trajectory of mental health comorbidities following gender affirmation. The current study examined mental healthcare and psychotropic medication utilization among TGD youth in a large healthcare administrative dataset, as compared to their siblings without a gender-related diagnosis, and explored mental health and psychotropic medication use in TGD adolescents following gender-affirming pharmaceutical care. We hypothesized that mental healthcare needs would be greater in TGD adolescents as compared to their siblings and that pharmaceutical affirmation would be associated with decreased treatment needs.

METHODS

We performed a retrospective cohort study examining mental healthcare utilization among TGD youth in the military healthcare system between October 2010 and September 2018 using the Military Healthcare Data Repository (MDR). The MDR includes records of all inpatient and outpatient care and outpatient prescriptions provided to military service members and retirees, and their family members domestically and abroad at military and civilian treatment facilities. The military provides no- to low-cost comprehensive care to these populations, including mental health and media (eg, pharmaceutical) care for gender dysphoria; active-duty members may also qualify for related surgical care. Access to care for gender dysphoria my differ geographically, however, the military attempts to address disparities through telemedicine and flying beneficiaries stationed overseas back to the United States, if needed, for specialized consultations. While generally reflective of the United States as a whole, the military tends to be more politically conservative, and more male, the proportion of African American individuals in the military is decreased, and the proportion of White individuals is increased as compared to the nation as a whole.[45]

TGD military dependent youth, <18 years of age at time of first contact, who received care in the military healthcare system were identified by 1 or more International Classification of Diseases (ICD) code (ICD-9 302.6, 302.85 302.50, 302.51, 302.52, 302.53, and ICD-10 F64.0, F64.1, F64.2, F64.8, F64.9, Z87.890) indicative of TGD status in their inpatient or outpatient record. This is a validated methodology.[8] ICD-9/10

codes are well-matched with clinical text notes in identification of TGD individuals.[46] We identified sibling controls in the MRD using the following criteria: shared a military sponsor (parent/guardian) with our TGD subjects; were <18 years old at their first encounter with the military health system during our study interval: and had no TGD diagnosis recorded, these siblings were considered cisgender controls, TDG youth and sibling were followed for the same time periods as care depended upon parental service.

We identified mental health care visits in the inpatient and outpatient care record by ICD-9/10 code using the Healthcare Cost and Utilization Project Clinical Classification Software system.[47] Mental health visits were sub-categorized by Healthcare Cost and Utilization Project categories for adjustment, anxiety, attention-deficit, conduct, developmental, mood, and cognitive disorders; disorders usually diagnosed in infancy or childhood (which includes autism), suicidal ideation/self-harm; alcohol use, substance use disorders; and miscellaneous mental health conditions (dissociative, eating and factitious disorders). The total number of mental health visits were counted overall, and by diagnosis sub-category; individuals were counted as having a given diagnosis if they had 1 or more visit for the diagnosis. Visits for TGD status, gender dysphoria, or mental health screening were not counted as mental health diagnoses. Children with one or more diagnosis for a mental health condition sub-category were categorized as having that mental health condition and having a mental health condition overall.

Psychotropic medications were identified by name in the outpatient pharmacy record, and included Bupropion, Selective Serotonin Reuptake Inhibitors (SSRI), Serotonin-Norepinephrine Reuptake Inhibitors (SNRI), other anti-depressants, sleep medication, benzodiazepines, antipsychotics, simulants, migraine medications, and lithium. Medications were classified by type and counted by day's supply. Gender affirming medications included puberty suppression (ie, implantable or injectable gonadotropin-releasing hormone agonists), masculinizing hormones and feminizing hormones, and were identified by name in the outpatient pharmacy record. Demographic data were extracted from the medical record and healthcare enrollment eligibility records; race/ethnicity data was not available.

Chi-squared analysis and Wilcoxon Rank sum test compared groups on demographics, logistic regression clustered by family compared groups on mental health diagnosis, and any psychotropic medication use overall and care/medication sub-category, and Poisson regression clustered by family compared mental healthcare visit rates and psychotropic medication days. Adjusted analyses controlled for sex assigned at birth, total healthcare contacts per year, age at study initiation, and parental rank. Parental rank was dichotomized as Junior Enlisted (enlisted ranks 1–4) vs more senior military ranks; Junior enlisted acted as a proxy for low income as Junior enlisted service member earn less than $35,000 a year.

Poisson regression clustered by individual compared mental healthcare visit rates and psychotropic medication days pre- and post- initiation of gender affirming pharmaceutical treatment, and logistic regression identified factors associated with decreased mental healthcare use and decreased psychotropic medications following initiation of gender-affirming medications. Adjusted models controlled for sex assigned at birth, total healthcare contacts per year, age at affirming medication initiation, type of initial gender affirming medication (puberty suppression vs gender-affirming hormones), and parental rank. Analyses were conducted using Stata Intercooled, version 13; P values <.05 was considered statistically significant. The study was reviewed and approved by the Uniformed Services University Institutional Review Board.

RESULTS

The research team identified a total of 3,754 TGD youths and 6,603 cisgender siblings who received Military Health System care between fiscal years 2010 and 2018. Both groups were tracked for a mean of 8.5 years. TGD youth were slightly older, less likely to be assigned male at birth, less likely to have Junior Enlisted parents, and utilized more outpatient healthcare overall as compared to their cisgender siblings (Table 1).

Mental Health Diagnosis

As compared to their cisgender siblings, TGD youth were more likely to have a mental health diagnosis and have a greater number of total mental health diagnoses (Table 1, 2). Looking at specific

Table 1. Demographics of included transgender and gender-diverse youth and their cisgender siblings

	TGD children N = 3,754	Cisgender siblings N = 6,603	Significance
Age at Study Initiation– Median [IQR]	10 [8-–13]	9 [4–14]	$P < .001$
Age at Study Completion– Median [IQR]	18 [16–21]	17 [11–21]	$P < .001$
Male Assigned Birth Sex	1,193 (31.8%)	3,312 (50.1%)	$P < .001$
Parent of Jr Enlisted Rank	1,524 (43.7%)	2960 (47.6%)	$P < .001$
Visits Per Year – Median [IQR]	18.7 [10.0–32.9]	9.5 [4.6–18.9]	$P < .001$
On psychotropic	2,820 (75.1%)	2,425 (37.7%)	$P < .001$
Years Tracked	8.5 [8.5–8.6]	8.5 [8.5–8.5]	$P < .001$
Median Mental Health Diagnoses	2 [1–4]	1 [0–2]	$P < .001$
Median Mental Health Visits Per Year	2.9 [0.8–7.0]	0.1 [0–2.0]	$P < .001$

TGD = transgender or gender-diverse.

Table 2. Mental health diagnoses and visits by transgender and gender-diverse status

	Mental health diagnosis			Visits per year		
	TGD Children N (%)	Cisgender Siblings N (%)	Adjusted* odds of mental health diagnosis or [95%CI]	TGD	Cisgender siblings	Adjusted* visit rate IRR [95%CI]
All Mental Health	3,352 (89.3)	3,308 (50.1)	5.45 [4.77–6.24]	5.5	3.1	2.22 [2.00–2.46]
Adjustment	1,687 (44.9)	1,191 (18.0)	1.09 [1.80–3.41]	0.74	0.29	2.49 [2.19–2.84]
Anxiety	1,908 (50.8)	1,216 (18.4)	3.30 [2.98–3.65]	0.77	0.28	2.49 [2.13–2.90]
ADHD	1,119 (29.8)	1,229 (18.6)	1.77 [1.59–1.97]	0.60	0.47	1.60 [1.37–1.88]
Cognitive	137 (3.7)	122 (1.9)	1.64 [1.26–2.14]	0.014	0.008	2.01 [1.39–2.89]
Developmental	189 (5.0)	429 (6.5)	1.11 [0.89–1.38]	0.10	0.30	0.97 [0.65–1.45]
First Diagnosed in Infancy	432 (11.5)	578 (8.8)	1.53 [1.30–1.79]	0.65	1.24	1.39 [0.86–2.26]
Impulse	60 (1.6)	45 (0.7)	2.18 [1.40–3.38]	0.013	0.009	1.55 [0.68–3.58]
Mood	2,413 (64.3)	1182 (18.9)	6.12 [5.51–6.80]	2.18	0.46	4.14 [3.64–4.71]
Personality	86 (2.3)	43 (0.7)	2.54 [1.71–3.78]	0.019	0.005	3.28 [1.53–7.00]
Psychotic	363 (9.7)	104 (1.6)	5.38 [4.20–6.88]	0.12	0.014	7.43 [4.72-–11.69]
Alcohol	57 (1.5)	66 (1.0)	1.25 [0.85–1.82]	0.011	0.010	0.93 [0.50–1.73]
Substance	237 (6.3)	209 (3.2)	1.61 [1.31–1.97]	0.053	0.032	1.77 [1.15–2.70]
Suicide	683 (18.2)	162 (2.5)	7.45 [6.11–9.08]	0.08	0.01	6.83 [5.03–9.26]
Miscellaneous	512 (13.6)	375 (5.7)	2.08 [1.77–2.45]	0.12	0.03	3.38 [2.20–5.18]

*Adjusted analysis, adjusts for sex assigned at birth, total healthcare contacts per year, age at study initiation, and parental rank.TGD = transgender or gender-diverse.

one-year periods, which is a more typical time period in which to access mental health diagnoses, TGD youth were more likely than their siblings to have a mental health diagnosis in a given year (eg, in 2010 [23.7% vs 13.9%, $P < .001$], 2015 [46.5% vs 18.8%, $P < .001$] and 2018 [42.7% vs 17.1%, $P < .001$]). The most common mental health diagnosis in TGD youth was mood/depressive disorder which impacted 64% of TGD youth at some point during the 8-year study period; followed by anxiety (51%) and adjustment disorders (44.9%; Table 2). For cisgender siblings, the most common mental health diagnoses were mood/depressive disorders (18.9%), ADHD (18.6%), and anxiety disorders (18.4%). After adjustment for age at study initiation, assigned sex at birth, parent rank, and number of outpatient visits per year, odds of having any mental health diagnosis were over 5 times higher in TGD youth as compared to their siblings (aOR = 5.45, 95% CI [4.77–6.24], $P < .001$). TGD youth were over 7 times as likely to have diagnosed suicidal ideation/self-harm (OR 7.45, 95%CI 6.11–9.08), over 6 times as likely to have a mood/depressive disorder (OR 6.12 95%CI [5.51–6.80]), over 5 times as likely to have a psychotic disorder (eg, schizophrenia) diagnosed (OR 5.38 95% CI [4.20–6.88]); and had similar odds of diagnosed developmental and alcohol use disorders (Table 2).

Mental Healthcare Use

TGD youth had an average of 5.5 mental healthcare visits per year over the course of the study as compared to 3.1 mental health visits per year for their cisgender siblings, and over twice as many visits in adjusted analysis (aIRR 2.22; 95% CI [2.00–2.46], $P < .001$). Mirroring diagnoses, mental healthcare visits for TGD youth were largely for mood/depressive, anxiety and adjustment disorders; however, care for cognitive, mood/depressive, personality, psychotic, and miscellaneous disorders, ADHD, substance use, and suicidal ideation/self-harm were all greater among TGD youth as compared to their siblings (Table 2). The most common diagnoses among siblings were disorders diagnosed in infancy and childhood, ADHD, and mood/depressive disorders. Care for development diagnoses, disorders usually diagnosed in infancy and childhood, impulse control disorder and alcohol use did not differ between the 2 groups (Table 2).

Psychotropic Medication Use

Over the full study period, 75% (2,820) of TGD youth were prescribed a psychotropic medication as compared to 38% (2,425) of their cisgender siblings ($P < .001$; Table 1). In adjusted analysis, TGD youth had over 2 and a half times as many medication days as their siblings (aIRR = 2.57; 95% CI [2.36–2.80], $P < .001$). SSRIs accounted for the most medication days in TGD youth, resulting in close to 3 times as many medication days for TGD youth; followed by stimulants, and antipsychotics. For siblings, stimulants accounted for the largest number of medication days followed by SSRIs, and anti-psychotics. In adjusted analyses TGD youth had over 3 times as many medication days for SNRIs, Lithium, anti-psychotics, and sleep medications as compared to their cisgender siblings (Table 3).

Impact of Gender Affirming Pharmaceutical Treatment

Of 3,754 included TGD youth, 963 (25.6%) initiated gender-affirming pharmaceutical treatment (puberty suppression or gender-affirming hormones) during the study period. The 963

Table 3. Psychotropic medication days by transgender and gender-diverse status

	Medication days per year		
	TGD children	Cisgender siblings	Adjusted[†] IRR [95% CI]
All Mental Health Meds	111.4	42.5	2.57 [2.36–2.80]
Wellbutrin	5.38	1.57	2.76 [2.12–3.60]
SSRI	37.25	11.18	2.96 [2.65–3.31]
SNRI	4.10	0.96	3.82 [2.64–5.54]
Other Antidepressant	7.93	2.50	3.01 [2.48–3.66]
Sleep Medications	5.82	1.61	3.28 [2.61–4.12]
Benzodiazepines	3.01	1.14	2.56 [1.85–3.56]
Anti-Psychotics	18.24	5.88	3.39 [2.83–4.07]
Stimulants	26.89	19.52	1.57 [1.39–1.77]
Migraine Medications*	0.92	0.42	1.69 [1.27–2.26]
Lithium	1.68	0.48	3.64 [2.02–6.55]

*Migraine Medications – Triptan.
[†]Adjusted analysis, adjusts for sex assigned at birth, total healthcare contacts per year, age at study initiation, and parental rank.TGD = transgender or gender-diverse.

pharmaceutically treated youth were tracked for a mean of 7.1 [IQR 5.5–7.8] years prior to pharmaceutical treatment initiation, and 1.5 [IQR 0.8–2.8] years following initiation of gender-affirming treatment. The median age of initiation of affirming medication was 18.2 [IQR 16.6–19.8] years, and the first gender affirming medications were: masculinizing hormones (61.4%, n = 591), feminizing hormones (28.7%, n = 276), and puberty suppression (10.0%, n = 96; Table 4). The median number of mental healthcare visits per years declined after starting gender affirming hormones (3.5 [IQR 1.2–7.5] vs 1.5 [0–7.8], Table 4). However, in adjusted Poisson regression analysis mental healthcare visits overall did not significantly change following gender-affirming pharmaceutical care (aIRR = 1.09, 95% CI [0.95–1.25], $P < .60$; 5.5 before vs 6.1 after) nor did care for most specific mental health diagnoses (Table 5). Of the youths who received gender-affirming pharmaceutical care, the majority (89%, n = 857) also used a psychotropic medication during the study period. Psychotropic medication use increased from mean of 120 days per year to a mean 212 days per year following gender affirming pharmaceutical care (aIRR = 1.67, 95%CI [1.46–1.91], $P < .001$]); medication use was increased in all classes explored except stimulants, migraine medications and lithium.

Factors Associated with Decreased Post-Affirming Mental Healthcare

Of youths receiving gender-affirming pharmaceutical care, 66.7% (642) had fewer mental healthcare visits following treatment. Decreased mental healthcare following gender-affirming care was associated with older age of medication initiation (aOR = 1.10, 95% CI [1.04–1.16]), and fewer overall visits per year over the study period (aOR = 0.99, 95%CI [0.99–0.99]), but was not associated with affirming medication type, sex assigned at birth, or parental rank. The median age of gender-affirming medication initiation of those with less mental healthcare use after initiation was 18.4 [IQR 17.0–19.8], and of those

Table 4. Demographics of 963 transgender and gender-diverse youth who initiated gender-affirming pharmaceutical treatment*

	TGD children N = 963		
	Before	After	*P*
Years Followed - Median	7.1 [5.6–7.9]	1.5 [0.7-2.7]	<.001
Mental Health Visits Per Year – Median [IQR]	3.5 [1.2–7.5]	1.5 [0-7.8]	<.001
Psychotropic Medication Days – Median [IQR]	69[17–157]	104[0-365]	.054
Fewer Mental Health Visits following Treatment	642 (66.7%)		
Fewer Medication Days Following Treatment	384 (44.8%)		
Age of First Affirming Medication	18.2 [16.6–19.8]		
First Medication Puberty Suppressant	96 (7.2%)		
First Medication Feminizing Hormone	276 (28.7%)		
First Medication Masculinizing Hormone	591 (61.4%)		
Male Sex Assigned at Birth	300 (31.2%)		
Parent of Jr Enlisted Rank	325 (33.8%)		
First Study Age – Median [IQR]	12 [10–14]		
Total Visits Per Year - Median[IQR]	48.9 [30.3–77.6]		

*Pharmaceutical treatment includes puberty suppression and gender-affirming hormonal therapy.TGD = transgender or gender-diverse.

Table 5. Transgender and gender-diverse youth mental healthcare and psychotropic medication use following initiation of gender affirming pharmaceutical treatment as compared to before initiation

Mental healthcare visits (N = 963)			
	Crude rate of visits per year		Adjusted* IRR [95% CI]
	Before	After	
All Mental Health Visits	5.50	6.10	1.04 [0.90-–1.20]
Adjustment	0.94	0.83	0.89 [0.67–-1.18]
Anxiety	0.98	1.04	1.07 [0.84–1.35]
ADHD	0.50	0.20	0.40 [0.27–-0.58]
Cognitive	0.02	0.02	0.83 [0.40–1.75]
Developmental	0.03	0.01	0.35 [0.16–0.78]
Infancy	0.37	0.53	1.02 [0.41–2.54]
Impulse	0.001	0.01	0.10 [0.02–0.53]
Mood	2.90	2.33	1.12 [0.94–1.35]
Personality	0.02	0.03	1.40 [0.44–-4.39]
Psychotic	0.13	0.16	0.99 [0.48–2.06]
Alcohol Abuse	0.13	0.06	0.66 [0.15–2.87]
Substance Abuse	0.05	0.12	1.39 [0.68–2.85]
Suicide	0.07	0.12	1.74 [1.18–2.56]
Miscellaneous	0.09	0.19	1.45 [0.56–3.60]

Medication days (N = 857)			
	Crude rate of medication days per year		Adjusted* IRR [95% CI]
	Before	After	
All Mental Health Meds	119.7	211.5	1.67 [1.46–1.91]
Wellbutrin	6.3	16.2	2.51 [2.71–3.69]
SSRI	44.8	73.9	1.72 [1.47–2.00]
SNRI	4.7	14.0	2.59 [1.52–4.38]
Other Antidepressant	9.2	18.9	1.61 [1.18–2.21]
Sleep Medications	6.4	16.2	2.23 [1.61–3.10]
Benzodiazepines	3.0	12.7	3.01 [1.95–4.65]
Anti-Psychotics	15.9	30.1	1.77 [1.34–2.35]
Stimulants	26.4	25.1	0.96 [0.72–1.26]
Migraine Medications	1.5	2.2	0.76 [0.37–1.53]
Lithium	1.3	2.3	1.11 [0.48–2.59]

*Adjusted analysis, adjusts for sex assigned at birth, total healthcare contacts per year, age at affirming medication initiation, and parental rank.

with more care was 17.9 [IQR 16.0–19.5]. Of included youth, 384 (44.8%) had decreased psychotropic medication prescription days following gender affirming pharmaceutical treatment. Decreased psychotropic medication use following gender affirming pharmaceutical treatment was associated with older age at time of affirming medication initiation (aOR = 1.09, 95% CI [1.03–1.16]) and male sex assigned at birth (aOR = 1.60 95% CI [1.18–2.17]).

DISCUSSION

Using a considerably larger population than previous studies, this research study found that TGD youth had greater mental healthcare use as compared to their cisgender siblings, with TGD youth more likely to have a mental health diagnosis, have multiple mental health diagnoses, use increased mental healthcare services, be prescribed a psychotropic medication, and use psychotropic medications for an increased number of days. For those TGD adolescents who initiated gender-affirming pharmaceutical care, mental healthcare and psychotropic medication needs were not reduced in the period following initiation after adjusting for confounders. Findings support previous research on a larger scale, control for family factors by comparing TGD youth to siblings, include psychotropic medication use as an additional mental health indicator, and document mental healthcare use rates as both an indicator of mental health severity and healthcare service need.

Over the 8.5-year course of the study's inclusion period, close to 90% of TGD youth had a mental health diagnosis, as compared to 50% of their cisgender siblings. For both TGD and cisgender youth, findings are higher than previously reported rates,[30,48] which likely relate to the study's extensive time period.

The median age of gender-affirming pharmaceutical treatment initiation was 18.2 years in this study which is substantially older than previous self- and parent-report studies of the initiation of medical and social transition (range 3−16 years).[32-37] Eighteen is the age at which youth can make their own medical choices; the fact that over half of included youth initiated gender-affirming pharmaceutical care after age 18 may suggest a lack of parental support or involvement in gender-affirming care among this study population. A lack of parental support may increase the need for new or ongoing mental health and/or psychotropic medication use which may also explain part of the increased rates of mental health diagnoses in our population. Yearly data from our study indicating that 23.7% to 46.5% of TGD youth had a mental health diagnosis in a given year is consistent with parental reports of mental health diagnoses in TGD youth at a given point in time.[29] Similarly, the yearly rate of mental health diagnoses among siblings (13.9%−18.8%) is comparable to published estimates.[48−50]

Findings of our study are consistent with adolescent and parent survey research, indicating that TGD youth have increased self- or parent-reported depression, suicide attempts/ideation, self-harm, substance use, and, emotional distress as compared to peers.[2,3,19,20,30] Results are also similar to large data research on adolescents and adults using clinic, healthcare provider, Medicaid, and veterans administration data which found increased mental health diagnoses in TGD individuals as compared to cisgender controls.[14,15,17,18,21,31] Our findings support current clinical recommendations to screen TGD youth for mental health concerns and address the underlying factors that increase risk in this population, and also suggest the importance of emphasizing mental health screening in future clinical recommendations.[51,52]

It is unclear why TGD youth were more likely to be diagnosed with psychotic conditions than their cisgender siblings, bur our findings are consistent with limited previous research in youth and adults.[31,52] Results may relate to lack of affirming care leading to depression with psychotic features.[53] Observed rates of increased provision of anti-psychotic medications among TGD youth may also be due to low dose prescriptions as an adjunct treatment for conditions such as severe depression and insomnia. The possible link between psychosis and TGD status warrants further exploration with well-validated psychiatric interviews.

While adjustment disorder, ADHD, cognitive, impulse control, personality, and miscellaneous diagnoses, substance use disorder, and conditions diagnosed in infancy or childhood (which includes autism) were significantly greater among TGD youth, differences were less pronounced. The finding of increased odds of conditions diagnosed in infancy and childhood, which includes autism, is consistent with previous research indicating increased odds of autism in TGD children and youth.[54,55] Developmental disorders and alcohol use disorders were not significantly increased among TGD youth.

Findings of increased psychotropic medication use (75% vs 38%) and medication days (111 vs 43 days per year) in TGD youth as compared to cisgender siblings are novel and corroborate prior studies indicating increased mental health needs in TGD youth.[2,3,14,15,17−21,30,31] Results are also consistent with findings that TGD adults were over 3 times as likely to use an antidepressant and/or anxiolytic,[41,61] but the current study is the first to examine psychotropic medications in youth, and include multiple medication classes. Consistent with increased care for anxiety, mood, and psychotic disorders; SSRIs, SNRIs, other antidepressants, Lithium, and anti-psychotics were all significantly increased in TGD youth (Table 3). TGD youth also had over 3 times as many sleep medication days as their cisgender siblings, results are consistent with research indicating a link between poor sleep duration/quality and depression/poor psychological well-being,[56,57] and suggests that screening for sleep concerns may be indicated in TGD care.

This study is among the first to analyze the associations of gender-affirming pharmaceutical treatment with mental health care patterns among TGD youth. Findings indicated that mental healthcare visits were not significantly changed and psychotropic medication use rose following gender-affirming pharmaceutical treatment after adjusting for potential confounders. Results are not consistent with adult and adolescent self-report survey research indicating improvements in mental health symptoms following gender-affirming care.[11,35−37,40,44] However, findings are consistent with one 10 year study which found visits for anxiety and mood disorders, and suicide attempt hospitalizations did not decrease following gender-affirming pharmaceutical care, but did decrease some following gender affirming surgery.[41,61] Findings that mental healthcare and psychotropic medications did not decrease after gender affirming care may be related to a number of factors. The median period following gender-affirming pharmaceutical care in the current study was relatively short (ie, 1.5 years), making it difficult to ascertain if the lack of a change in care patterns was related to continuing mental health problems, or represents the delivery of responsible mental healthcare that maintains a therapeutic relationship through a substantial life transition. Similarly, the period before initiation of gender affirming care was 7.1 years, making it possible that mental healthcare during the earlier portion of this period is not reflective of mental healthcare use patterns of youth with gender dysphoria, artificially deflating the rate of care in the period before gender affirming pharmaceutical care. Patients also age during the pretreatment period and mental health utilization may increase over time irrespective of gender affirming care.

Also, the sample in this study may differ from samples previously recruited from specialty transgender clinics. Military connected families are generally more conservative,[58] which may relate to the relatively low percentage receiving puberty blockers, and relatively older age of starting gender affirming pharmaceutical care. Military connected children and youth also have free mental healthcare and psychiatric medications through the age of 23, which may lower barriers to continued engagement in treatment of mental health conditions after gender transition. This

care would allow patients and clinicians to thoroughly address and treat all identified issues irrespective of gender-affirming treatment status, and maintain engagement in ongoing care even as symptoms begin to remit.[59]

The impact of access to high quality no-or low-cost mental healthcare available to the study population may impact mental healthcare and psychotropic medication trends, making results potentially less generalizable for adolescents and young adults in the United States. Although adults in the United States report some of the highest rates of mental health conditions, access to mental healthcare in the United States is reduced as compared to other high-income countries.[60]

Older age and fewer yearly visits were associated with decreased post-affirmation mental healthcare, and older age and male sex at birth were associated with decreased post-affirmation psychotropic medication prescriptions. Findings that older age was associated with decreased mental health and psychotropic care may suggest that parents were involved in scheduling and seeking mental healthcare and psychotropic medications for their younger children. Conversely, older youth who make their own medication decisions may have difficulties in scheduling care, or decide to reduce care they do not deem necessary. Alternatively, patients with higher levels of distress or engagement with mental health providers may be more likely to have parents that acknowledge their distress and consent for treatment.

Strengths of this study include the very large sample size, inclusion of data on psychiatric diagnoses, mental healthcare visits, and psychotropic medication use to assess mental health disparities, the extensive study period, the assessment of mental health care utilization following gender affirmation treatment, and the use of sibling controls. A sibling study group controls for household healthcare use, threshold for accessing mental healthcare, and gender socialization experience, but does not account for all differences between individuals. This study is limited by the use of healthcare data in the form of ICD-9/10 codes which cannot indicate the severity of diagnoses or the full breadth of complex TGD identities; however, the use of multiple indicators of mental health burden does mitigate this concern. The study is also limited by the short duration of care following gender-affirming pharmaceutical treatment, which may be insufficient to observe any clinically significant change. We were also unable to control for differing, regional, family level, and care provider acceptance, however within the military access to specialists can occur when requested. We also didn't distinguish puberty suppression from testosterone/estrogens as we were interested in pharmaceuticals as an indicator of treatment progression; however, it is possible that there are differences in outcomes for the 2 groups. Furthermore, the effect of affirming medical care may be confounded by increasing mental health disparities as TGD youth age (eg, due to increasing minority stress). Finally, results may have limited generalizability as military dependent youth face additional stressors, such as multiple moves and parental deployment, and benefit from high-quality free military healthcare until age 21 (or age 23 if in college), thereby potentially affecting post-affirming healthcare use patterns. The military population included in this study is likely substantially different than previous research which generally recruited youth from specialized TGD care clinics (which signals parental and family support). Available data did not allow for study of important intersections between cultural and ethnic factors which may predict outcomes such as healthcare utilization. While many adolescents included in this study received care for GD prior to age 18, many did not, and pharmaceutical affirmation was initiated after age 18 for over 50% of include youth. Therefore, results of this study may be more representative of the national population of TGD youth, some of whom receive parental support and some of whom self-report a lack of parental and family support, than the specialty clinic population used in many previous studies.

CONCLUSIONS

TGD youth have considerably greater mental health diagnoses, care, and psychotropic medication use across a range of diagnoses, as compared to their cisgender siblings. Results strongly support clinical recommendations for screening of mental health conditions in TGD youth and availability of healthcare for those in need. Additional research is needed to determine the long-term impact of gender-affirming care on psychiatric co-morbidities among TGD youth and young adults. While the need for mental health treatment appears to persist after the initiation of gender-affirming pharmaceutical treatment, longer term follow-up and care-specific analysis is needed to accurately understand changing care needs over time. Results may have policy implications as some states are currently considering limiting gender affirming care to adolescents.

DISCLAIMER

The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the Uniformed Services University, the U.S. Air Force, the U.S. Navy, the U.S. Department of Defense, or the U.S. Government.

Corresponding Author: Elizabeth Hisle-Gorman, MSW, PhD, F. Edward Hébert School of Medicine - "America's Medical School", Uniformed Services University, Bethesda, MD 20814, USA. Tel: 301-221-7546/301-295-9726; Fax: 301-295-3898; E-mail: elizabeth.hisle-gorman.ctr@usuhs.edu

Conflict of Interest: The authors report no competing interests.

Funding: None.

Parts of this research were presented at the annual meeting for the Society of Adolescent Health and Medicine (virtual) in March 2020.

STATEMENT OF AUTHORSHIP

Dr. Hisle-Gorman was responsible for conception and design of the study, analysis, and interpretation of results, drafting the article, and final approval of the paper.

Dr. Schvey was responsible for obtaining permission for data acquisition, assisting with the analysis plan, and revising the manuscript for intellectual content, and final approval of the paper.

Dr. Adirim was responsible for assisting with interpretation of results, revising the paper for critical content and final approval of the paper.

Dr. Rayne was responsible for helping to draft the paper, and approval of the final paper

Ms. Susi was responsible for acquisition of the data, and initial cleaning and analysis, revisiting the paper critically and approval of the final paper to be submitted.

Dr. Roberts was responsible for assisting with conception and design of the study, revising it critically for important intellectual content and approval of the final version to be submitted.

Dr. Klein was responsible for assisting with conception and design of the study, interpretation of results, revising the paper critically for important intellectual content, and final approval of the paper.

REFERENCES

1. Rafferty J, Committee On Psychosocial Aspects Of Child and Family Health Committee on Adolescence and Section on Lesbian, Gay, Bisexual and Transgender Health and Wellness. Ensuring comprehensive care and support for transgender and gender-diverse children and adolescents. **Pediatrics 2018;142 (4):1–16.**
2. Clark TC, Lucassen MF, Bullen P, et al. The health and well-being of transgender high school students: results from the New Zealand adolescent health survey (Youth'12). **J Adolesc Health 2014;55:93–99.**
3. Eisenberg ME, Gower AL, McMorris BJ. Risk and protective factors in the lives of transgender/gender nonconforming adolescents. **J Adolesc Health 2017;61:521–526.**

[4]. Herman JL, Flores AR, Brown TNT, et al. Age of Individuals who Identify as Transgender in the United States. Los Angeles, CA: UCLA School of Law; 2017.

5. Shields JP, Cohen R, Glassman JR, et al. Estimating population size and demographic characteristics of lesbian, gay, bisexual, and transgender youth in middle school. **J Adolesc Health 2013;52:248–250.**
6. Klein DA, Roberts TA, Adirim TA, et al. Transgender children and adolescents receiving care in the US military health care system. **JAMA Pediatr 2019;173:491–492.**
7. Zucker KJ. Epidemiology of gender dysphoria and transgender identity. **Sex Health 2017;14:404–411.**
8. Ewald ER, Guerino P, Dragon C, et al. Identifying medicare beneficiaries accessing transgender-related care in the era of ICD-10. **LGBT Health 2019;6:166–173.**
9. Handler T, Hojilla JC, Varghese R, et al. Trends in referrals to a pediatric transgender clinic. **Pediatrics 2019;144(5):1–9.**
10. Arnoldussen M, Steensma TD, Popma A, et al. Re-evaluation of the Dutch approach: are recently referred transgender youth different compared to earlier referrals? **Eur Child Adolesc Psychiatry 2020;29:803–811.**
11. Connolly MD, Zervos MJ, Barone 2nd CJ, et al. The mental health of transgender youth: advances in understanding. **J Adolesc Health 2016;59:489–495.**
12. Stall R, Matthews DD, Friedman MR, et al. The continuing development of health disparities research on lesbian, gay, bisexual, and transgender individuals. **Am J Public Health 2016;106:787–789.**
13. Adelson SL, American Academy of C, Adolescent Psychiatry Committee on Quality I. Practice parameter on gay, lesbian, or bisexual sexual orientation, gender nonconformity, and gender discordance in children and adolescents. **J Am Acad Child Adolesc Psychiatry 2012;51:957–974.**
14. Downing J, Conron K, Herman JL. Transgender and cisgender US veterans have few health differences. **Health Aff (Millwood) 2018;37:1160–1168.**
15. Dragon CN, Guerino P, Ewald E, et al. Transgender medicare beneficiaries and chronic conditions: exploring fee-for-service claims data. **LGBT Health 2017;4:404–411.**
16. Progovac AM, Cook BL, Mullin BO, et al. Identifying gender minority patients' health and health care needs in administrative claims data. **Health Aff (Millwood) 2018;37:413–420.**
17. Blosnich JR, Brown GR, Shipherd Phd JC, et al. Prevalence of gender identity disorder and suicide risk among transgender veterans utilizing veterans health administration care. **Am J Public Health 2013;103:e27–e32.**
18. Blosnich JR, Brown GR, Wojcio S, et al. Mortality among veterans with transgender-related diagnoses in the veterans health administration, FY2000-2009. **LGBT Health 2014;1: 269–276.**
19. Lytle MC, Blosnich JR, Kamen C. The association of multiple identities with self-directed violence and depression among transgender individuals. **Suicide Life Threat Behav 2016; 46:535–544.**
20. Veale JF, Watson RJ, Peter T, et al. Mental health disparities among Canadian transgender youth. **J Adolesc Health 2017; 60:44–49.**
21. Reisner SL, Vetters R, Leclerc M, et al. Mental health of transgender youth in care at an adolescent urban community health center: a matched retrospective cohort study. **J Adolesc Health 2015;56:274–279.**
22. Blosnich JR, Marsiglio MC, Gao S, et al. Mental health of transgender veterans in US states with and without discrimination and hate crime legal protection. **Am J Public Health 2016;106:534–540.**

23. Bukowski LA, Blosnich J, Shipherd JC, et al. Exploring ural disparities in medical diagnoses among veterans with transgender-related diagnoses utilizing veterans health administration care. **Med Care 2017;55 Suppl 9(Suppl 2): S97–S103.**
24. Johns MM, Lowry R, Andrzejewski J, et al. Transgender identity and experiences of violence victimization, substance use, suicide risk, and sexual risk behaviors among high school students - 19 states and large urban school districts, 2017. **MMWR Morb Mortal Wkly Rep 2019;68:67–71.**
25. Olson J, Schrager SM, Belzer M, et al. Baseline physiologic and psychosocial characteristics of transgender youth seeking care for gender dysphoria. **J Adolesc Health 2015;57:374–380.**
26. Blosnich JR, Lehavot K, Glass JE, et al. Differences in alcohol use and alcohol-related health care among transgender and nontransgender adults: findings from the 2014 behavioral risk factor surveillance system. **J Stud Alcohol Drugs 2017; 78:861–866.**
27. Coulter RW, Blosnich JR, Bukowski LA, et al. Differences in alcohol use and alcohol-related problems between transgender- and nontransgender-identified young adults. **Drug Alcohol Depend 2015;154:251–259.**
28. Menino DD, Katz-Wise SL, Vetters R, et al. Associations between the length of time from transgender identity recognition to hormone initiation and smoking among transgender youth and young adults. **Transgend Health 2018;3:82–87.**
29. Wang Y, Yu H, Yang Y, et al. Mental health status of cisgender and gender-diverse secondary school students in China. **JAMA Netw Open 2020;3:e2022796.**
30. de Vries AL, Doreleijers TA, Steensma TD, et al. Psychiatric comorbidity in gender dysphoric adolescents. **J Child Psychol Psychiatry 2011;52:1195–1202.**
31. Becerra-Culqui TA, Liu Y, Nash R, et al. Mental health of transgender and gender nonconforming youth compared with their peers. **Pediatrics 2018;141(5):1–13.**
32. Olson KR, Durwood L, DeMeules M, et al. Mental health of transgender children who are supported in their identities. **Pediatrics 2016;137:e20153223.**
33. Durwood L, McLaughlin KA, Olson KR. Mental health and self-worth in socially transitioned transgender youth. **J Am Acad Child Adolesc Psychiatry 2017;56:116–123.e2 116-123 e112.**
34. van der Miesen AIR, Steensma TD, de Vries ALC, et al. Psychological functioning in transgender adolescents before and after gender-affirmative care compared with cisgender general population peers. **J Adolesc Health 2020;66:699–704.**
35. de Vries AL, McGuire JK, Steensma TD, et al. Young adult psychological outcome after puberty suppression and gender reassignment. **Pediatrics 2014;134:696–704.**
36. Chew D, Anderson J, Williams K, et al. Hormonal treatment in young people with gender dysphoria: a systematic review. **Pediatrics 2018;141(4):1–20.**
37. de Vries AL, Steensma TD, Doreleijers TA, et al. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. **J Sex Med 2011;8:2276–2283.**
38. Heylens G, Verroken C, De Cock S, et al. Effects of different steps in gender reassignment therapy on psychopathology: a prospective study of persons with a gender identity disorder. **J Sex Med 2014;11:119–126.**
39. Colizzi M, Costa R, Todarello O. Transsexual patients' psychiatric comorbidity and positive effect of cross-sex hormonal treatment on mental health: results from a longitudinal study. **Psychoneuroendocrinology 2014;39:65–73.**
40. Tucker RP, Testa RJ, Simpson TL, et al. Hormone therapy, gender affirmation surgery, and their association with recent suicidal ideation and depression symptoms in transgender veterans. **Psychol Med 2018;48:2329–2336.**
41. Branstrom R, Pachankis JE. Reduction in mental health treatment utilization among transgender individuals after gender-affirming surgeries: a total population study. **Am J Psychiatry 2020;177:727–734 appiajp201919010080.**
42. Kuper LE, Stewart S, Preston S, et al. Body dissatisfaction and mental health outcomes of youth on gender-affirming hormone therapy. **Pediatrics 2020;145(4):1–11.**
43. Turban JL, King D, Carswell JM, et al. Pubertal suppression for transgender youth and risk of suicidal ideation. **Pediatrics 2020;145(2):1–10.**
44. Murad MH, Elamin MB, Garcia MZ, et al. Hormonal therapy and sex reassignment: a systematic review and meta-analysis of quality of life and psychosocial outcomes. **Clin Endocrinol (Oxf) 2010;72:214–231.**
45. Branstrom R, Pachankis J. Reduction in mental health treatment utilization among transgender individuals after gender-affirming surgeries: A total population study - correction to branstrom and pachankis. **Am J Psychiatry 2020;177(8):734.**
46. Military One Source. 2019 Demographics Profile of the Military Community. 2019 Demographics Profile of the Military Community. Washington, DC: Office of the Deputy Assistant Secretary of Defense for Military Community and Family Policy; 2019 Department of Defense (DoD), Office of the Deputy Assistant Secretary fo Defense for Military Community and Family Policy.
47. Blosnich JR, Cashy J, Gordon AJ, et al. Using clinician text notes in electronic medical record data to validate transgender-related diagnosis codes. **J Am Med Inform Assoc 2018;25:905–908.**
48. HCUP Databases. Healthcare cost and utilization project (HCUP). Agency for healthcare reserach and quality published 2006-2009.
49. Hisle-Gorman E, Susi A, Gorman GH. Mental health trends in military pediatrics. **Psychiatr Serv 2019;70: 657–664.**
50. Perou R, Bitsko RH, Blumberg SJ, et al. Mental health surveillance among children—United States, 2005-2011. **MMWR Suppl 2013;62:1–35.**
51. Merikangas KR, He JP, Brody D. Prevalence and treatment of mental disorders among US children in the 2001-2004 NHANES. **Pediatrics 2010;125:75–81.**

52. Society for Adolescent Health and Medicine. Recommendations for promoting the health and well-being of lesbian, gay, bisexual, and transgender adolescents: a position paper of the society for adolescent health and medicine. **J Adolesc Health 2013;52:506–510.**
53. Coleman E, Bockting W, Botzer M, et al. Standards of Care for the Health of Transsexual, Transgender, and Gender-Nonconforming People, Version 7. International Journal of Transgenderism; 2012. p. 165–232.
54. Smith WB, National LGBT Health Education Center A Program of the Fenway Institute. Caring for Transgender People with Severe Mental Illness. Caring for Transgender People with Severe Mental Illness. Boston, MA: The Fenway Institute; 2018.
55. Hisle-Gorman E, Landis CA, Susi A, et al. Gender dysphoria in children with Autism spectrum disorder. **LGBT Health 2019;6:95–100.**
56. Warrier V, Greenberg DM, Weir E, et al. Elevated rates of autism, other neurodevelopmental and psychiatric diagnoses, and autistic traits in transgender and gender-diverse individuals. **Nat Commun 2020;11:3959.**
57. Zhai K, Gao X, Wang G. The role of sleep quality in the psychological well-being of final year undergraduate students in China. **Int J Environ Res Public Health 2018;15 (12):2881–2893.**
58. Zhai L, Zhang H, Zhang D. Sleep duration and depression among adults: a meta-analysis of prospective studies. **Depress Anxiety 2015;32:664–670.**
59. Maniam S. U.S. veterans are generally supportive of Trump. **Pew Research Center 2017. https://www.pewresearch.org/fact-tank/2017/05/26/u-s-veterans-are-generally-supportive-of-trump/. Accessed June 29, 2021.**
60. Fraser L, Knudson G. Past and future challenges associated with standards of care for gender transitioning clients. **Psychiatr Clin North Am 2017;40:15–27.**
61. Tikkanen R, Fields K, Williams II R, et al. Mental health conditions and substance use: comparing U.S. needs and treatment capacity with those in other high-income countries. The commonwealth fund. Data brief web site. Available at: https://www.commonwealthfund.org/sites/default/files/2020-05/Tikkanen_-mental_hlt_intl_comparison_db.pdf. Accessed April 2, 2021.

ENDO REPORTS FOURTH-QUARTER AND FULL-YEAR 2021 FINANCIAL RESULTS

DUBLIN, February 28, 2022 -- Endo International plc (NASDAQ: ENDP) today reported financial results for the fourth-quarter and full-year ended December 31, 2021 and introduced first-quarter 2022 financial guidance.

"Endo's solid operational and financial performance in the fourth-quarter ended a year of strong performance across all of our business segments," said Blaise Coleman, President and Chief Executive Officer at Endo. "As we look forward to 2022 and begin to transition through the VASOSTRICT® loss of exclusivity and the ongoing COVID-19 driven market conditions negatively impacting specialty product office-based procedures, we remain focused on advancing our strategic priorities supported by strong commercial execution to maximize XIAFLEX®, continuing to establish QWO® as a cornerstone treatment for cellulite and investing to advance our pipeline in our core areas of growth."

FOURTH-QUARTER FINANCIAL PERFORMANCE

(in thousands, except per share amounts)

	Three Months Ended December 31,			Year Ended December 31,		
	2021	2020	Change	2021	2020	Change
Total Revenues, Net	$ 789,429	$ 760,221	4 %	$ 2,993,206	$ 2,903,074	3 %
Reported (Loss) Income from Continuing Operations	$ (556,667)	$ 141,247	NM	$ (569,081)	$ 247,464	NM
Reported Diluted Weighted Average Shares	233,681	234,474	— %	232,785	233,653	— %
Reported Diluted Net (Loss) Income per Share from Continuing Operations	$ (2.38)	$ 0.60	NM	$ (2.44)	$ 1.06	NM
Reported Net (Loss) Income	$ (562,062)	$ 119,343	NM	$ (613,245)	$ 183,944	NM
Adjusted Income from Continuing Operations (2)	$ 200,034	$ 175,995	14 %	$ 716,349	$ 670,370	7 %
Adjusted Diluted Weighted Average Shares (1)(2)	237,045	234,474	1 %	236,665	233,653	1 %
Adjusted Diluted Net Income per Share from Continuing Operations (2)	$ 0.84	$ 0.75	12 %	$ 3.03	$ 2.87	6 %
Adjusted EBITDA (2)	$ 386,524	$ 351,635	10 %	$ 1,480,822	$ 1,395,942	6 %

(1) Reported Diluted Net (Loss) Income per Share from Continuing Operations is computed based on weighted average shares outstanding and, if there is income from continuing operations during the period, the dilutive impact of ordinary share equivalents outstanding during the period. In the case of Adjusted Diluted Weighted Average Shares, Adjusted Income from Continuing Operations is used in determining whether to include such dilutive impact.

(2) The information presented in the table above includes non-GAAP financial measures such as "Adjusted Income from Continuing Operations," "Adjusted Diluted Weighted Average Shares," "Adjusted Diluted Net Income per Share from Continuing Operations" and "Adjusted EBITDA." Refer to the "Supplemental Financial Information" section below for reconciliations of certain non-GAAP financial measures to the most directly comparable GAAP financial measures.

CONSOLIDATED RESULTS

Total revenues were $789 million in fourth-quarter 2021, an increase of 4% compared to $760 million during the same period in 2020. This increase was attributable to increased revenues across our Branded, Generic and International Pharmaceuticals segments, partially offset by decreased revenues from our Sterile Injectables segment.

Reported loss from continuing operations in fourth-quarter 2021 was $557 million compared to reported income from continuing operations of $141 million during the same period in 2020. This decrease was primarily due to increased asset impairment charges, opioid settlement and litigation-related costs, and income tax expense, due to a non-cash income tax benefit recorded in the fourth-quarter 2020, partially offset by increased revenues and favorable changes in product mix. Reported diluted net loss per share from continuing operations in fourth-quarter 2021 was $2.38 compared to reported diluted net income per share from continuing operations in fourth-quarter 2020 of $0.60.

Adjusted income from continuing operations in fourth-quarter 2021 was $200 million compared to $176 million in fourth-quarter 2020. The result was attributable to increased revenues and favorable changes in product mix. Adjusted diluted net income per share from continuing operations in fourth-quarter 2021 was $0.84 compared to $0.75 in fourth-quarter 2020.

BRANDED PHARMACEUTICALS SEGMENT

Fourth-quarter 2021 Branded Pharmaceuticals segment revenues were $228 million, an increase of 2% compared to $225 million during fourth-quarter 2020.

Despite increasing COVID-19 driven pressures during the fourth-quarter 2021, Specialty Products revenues increased 4% to $161 million in fourth-quarter 2021 compared to $154 million in fourth-quarter 2020. XIAFLEX® revenues increased 14% to $120 million compared to $105 million in fourth-quarter 2020, driven by increased net price and improving patient demand compared to the prior year. Established Products revenues decreased 5% to $67 million in fourth-quarter 2021 compared to $71 million in fourth-quarter 2020 due to ongoing competitive pressures in the portfolio.

STERILE INJECTABLES SEGMENT

Fourth-quarter 2021 Sterile Injectables segment revenues were $319 million, a decrease of 4% compared to $332 million during fourth-quarter 2020. This decrease was primarily attributable to competitive pressure on certain products, which was partially offset by higher VASOSTRICT® revenues primarily due to hospitalizations associated with COVID-19.

GENERIC PHARMACEUTICALS SEGMENT

Fourth-quarter 2021 Generic Pharmaceuticals segment revenues were $218 million, an increase of 21% compared to $180 million during fourth-quarter 2020. This increase was primarily attributable to additional revenues from 2021 product launches, including lubiprostone capsules, the first authorized generic of Amitiza®, and varenicline tablets, the only FDA-approved generic version of Chantix®, partially offset by competitive pressure on certain other generic products.

INTERNATIONAL PHARMACEUTICALS SEGMENT

Fourth-quarter 2021 International Pharmaceuticals segment revenues were $24 million compared to $23 million during fourth-quarter 2020.

FIRST-QUARTER 2022 FINANCIAL GUIDANCE

Due to uncertainties in certain key assumptions including the timing and impact of VASOSTRICT® generic competition and the rate and extent to which the market for specialty product office-based procedures recovers from the current COVID-19 driven challenges, the Company is only providing financial guidance for the first quarter ending March 31, 2022 at this time. These statements are forward-looking, and actual results may differ materially from Endo's expectations, as further discussed below under the heading "Cautionary Note Regarding Forward-Looking Statements."

	First-Quarter 2022
Total Revenues, Net	$595 - $635M
Adjusted EBITDA	$240 - $260M
Adjusted Diluted Net Income per Share from Continuing Operations	$0.35 - $0.45
Assumptions:	
Adjusted Gross Margin	~71.5%
Adjusted Operating Expenses as a Percentage of Total Revenues, Net	~33.0%
Adjusted Interest Expense	~$140M
Adjusted Effective Tax Rate	~$1.0%
Adjusted Diluted Weighted Average Shares	~238M

BALANCE SHEET, LIQUIDITY AND OTHER UPDATES

As of December 31, 2021, the Company had approximately $1.5 billion in unrestricted cash; $8.2 billion of debt; and a net debt to adjusted EBITDA ratio of 4.6.

Fourth-quarter 2021 net cash used in operating activities was $50 million compared to $108 million provided by operating activities during the fourth-quarter 2020. This change was primarily due to higher payments in 2021 related to interest, opioid-related and other legal settlements and expenses and continuity and separation benefits, cost reduction and strategic review initiatives, partially offset by an increase in adjusted income from continuing operations.

Additionally, during the fourth-quarter 2021, the Company completed the previously announced sales of its manufacturing sites in Chestnut Ridge, New York and Irvine, California. The exit of these sites was included in a series of business transformation initiatives that the Company announced in late 2020, including further optimization of its generic retail business cost structure.

CONFERENCE CALL INFORMATION

Endo will conduct a conference call with financial analysts to discuss this press release tomorrow, March 1, 2022, at 7:30 a.m. ET. The dial-in number to access the call is U.S./Canada (866) 497-0462, International (678) 509-7598, and the passcode is 4272796. Please dial in 10 minutes prior to the scheduled start time.

A replay of the call will be available from March 1, 2022 at 10:30 a.m. ET until 9:30 a.m. ET on March 8, 2022 by dialing U.S./Canada (855) 859-2056 International (404) 537-3406, and entering the passcode 4272796.

A simultaneous webcast of the call can be accessed by visiting https://investor.endo.com/events-and-presentations. In addition, a replay of the webcast will be available on the Company website for one year following the event.

Chantix® is a registered trademark of Pfizer Inc.
Amitiza® is a registered trademark of Mallinckrodt plc.

FINANCIAL SCHEDULES

The following table presents Endo's unaudited Total revenues, net for the three months and years ended December 31, 2021 and 2020 (dollars in thousands):

	Three Months Ended December 31,		Percent Growth	Year Ended December 31,		Percent Growth
	2021	2020		2021	2020	
Branded Pharmaceuticals:						
Specialty Products:						
XIAFLEX®	$ 120,078	$ 105,212	14 %	$ 432,344	$ 316,234	37 %
SUPPRELIN® LA	28,709	24,838	16 %	114,374	88,182	30 %
Other Specialty (1)	12,025	23,867	(50)%	86,432	92,662	(7)%
Total Specialty Products	$ 160,812	$ 153,917	4 %	$ 633,150	$ 497,078	27 %
Established Products:						
PERCOCET®	$ 25,093	$ 27,323	(8)%	$ 103,788	$ 110,112	(6)%
TESTOPEL®	11,322	8,357	35 %	43,636	35,234	24 %
Other Established (2)	30,738	34,907	(12)%	113,043	139,356	(19)%
Total Established Products	$ 67,153	$ 70,587	(5)%	$ 260,467	$ 284,702	(9)%
Total Branded Pharmaceuticals (3)	$ 227,965	$ 224,504	2 %	$ 893,617	$ 781,780	14 %
Sterile Injectables:						
VASOSTRICT®	$ 224,971	$ 213,116	6 %	$ 901,735	$ 785,646	15 %
ADRENALIN®	36,494	31,739	15 %	124,630	152,074	(18)%
Other Sterile Injectables (4)	57,634	86,995	(34)%	239,732	301,127	(20)%
Total Sterile Injectables (3)	$ 319,099	$ 331,850	(4)%	$ 1,266,097	$ 1,238,847	2 %
Total Generic Pharmaceuticals	$ 218,135	$ 180,440	21 %	$ 740,586	$ 783,110	(5)%
Total International Pharmaceuticals	$ 24,230	$ 23,427	3 %	$ 92,906	$ 99,337	(6)%
Total revenues, net	$ 789,429	$ 760,221	4 %	$ 2,993,206	$ 2,903,074	3 %

(1) Products included within Other Specialty include NASCOBAL® Nasal Spray, AVEED® and QWO®.

(2) Products included within Other Established include, but are not limited to, EDEX® and LIDODERM®.

(3) Individual products presented above represent the top two performing products in each product category for the year ended December 31, 2021 and/or any product having revenues in excess of $25 million during any quarterly period in 2021 or 2020.

(4) Products included within Other Sterile Injectables include ertapenem for injection, APLISOL® and others.

The following table presents unaudited Condensed Consolidated Statement of Operations data for the three months and years ended December 31, 2021 and 2020 (in thousands, except per share data):

	Three Months Ended December 31,		Year Ended December 31,	
	2021	2020	2021	2020
TOTAL REVENUES, NET	$ 789,429	$ 760,221	$ 2,993,206	$ 2,903,074
COSTS AND EXPENSES:				
Cost of revenues	311,223	369,539	1,221,064	1,442,511
Selling, general and administrative	250,103	176,221	861,760	698,506
Research and development	58,536	64,737	148,560	158,902
Litigation-related and other contingencies, net	226,168	4,889	345,495	(19,049)
Asset impairment charges	364,584	14,147	414,977	120,344
Acquisition-related and integration items, net	(2,022)	(551)	(8,379)	16,549
Interest expense, net	143,501	135,250	562,353	532,939
Loss on extinguishment of debt	—	—	13,753	—
Other (income) expense, net	(15,103)	4,208	(19,774)	(21,110)
LOSS FROM CONTINUING OPERATIONS BEFORE INCOME TAX	$ (547,561)	$ (8,219)	$ (546,603)	$ (26,518)
INCOME TAX EXPENSE (BENEFIT)	9,106	(149,466)	22,478	(273,982)
(LOSS) INCOME FROM CONTINUING OPERATIONS	$ (556,667)	$ 141,247	$ (569,081)	$ 247,464
DISCONTINUED OPERATIONS, NET OF TAX	(5,395)	(21,904)	(44,164)	(63,520)
NET (LOSS) INCOME	$ (562,062)	$ 119,343	$ (613,245)	$ 183,944
NET (LOSS) INCOME PER SHARE—BASIC:				
Continuing operations	$ (2.38)	$ 0.61	$ (2.44)	$ 1.08
Discontinued operations	(0.03)	(0.09)	(0.19)	(0.28)
Basic	$ (2.41)	$ 0.52	$ (2.63)	$ 0.80
NET (LOSS) INCOME PER SHARE—DILUTED:				
Continuing operations	$ (2.38)	$ 0.60	$ (2.44)	$ 1.06
Discontinued operations	(0.03)	(0.09)	(0.19)	(0.27)
Diluted	$ (2.41)	$ 0.51	$ (2.63)	$ 0.79
WEIGHTED AVERAGE SHARES:				
Basic	233,681	230,301	232,785	229,314
Diluted	233,681	234,474	232,785	233,653

The following table presents unaudited Condensed Consolidated Balance Sheet data at December 31, 2021 and December 31, 2020 (in thousands):

	December 31, 2021	December 31, 2020
ASSETS		
CURRENT ASSETS:		
Cash and cash equivalents	$ 1,507,196	$ 1,213,437
Restricted cash and cash equivalents	124,114	171,563
Accounts receivable	592,019	511,262
Inventories, net	283,552	352,260
Other current assets	207,705	164,736
Total current assets	$ 2,714,586	$ 2,413,258
TOTAL NON-CURRENT ASSETS	6,052,829	6,851,379
TOTAL ASSETS	$ 8,767,415	$ 9,264,637
LIABILITIES AND SHAREHOLDERS' DEFICIT		
CURRENT LIABILITIES:		
Accounts payable and accrued expenses, including legal settlement accruals	$ 1,417,892	$ 1,208,061
Other current liabilities	212,070	45,763
Total current liabilities	$ 1,629,962	$ 1,253,824
LONG-TERM DEBT, LESS CURRENT PORTION, NET	8,048,980	8,280,578
OTHER LIABILITIES	332,459	378,174
SHAREHOLDERS' DEFICIT	(1,243,986)	(647,939)
TOTAL LIABILITIES AND SHAREHOLDERS' DEFICIT	$ 8,767,415	$ 9,264,637

The following table presents unaudited Condensed Consolidated Statement of Cash Flow data for the years ended December 31, 2021 and 2020 (in thousands):

	Year Ended December 31,	
	2021	2020
OPERATING ACTIVITIES:		
Net (loss) income	$ (613,245)	$ 183,944
Adjustments to reconcile Net (loss) income to Net cash provided by operating activities:		
Depreciation and amortization	457,098	518,807
Asset impairment charges	414,977	120,344
Other, including cash payments to claimants from Qualified Settlement Funds	152,220	(425,703)
Net cash provided by operating activities	$ 411,050	$ 397,392
INVESTING ACTIVITIES:		
Capital expenditures, excluding capitalized interest	$ (77,929)	$ (69,971)
Acquisitions, including in-process research and development, net of cash and restricted cash acquired	(5,000)	(649,504)
Proceeds from sales and maturities of investments	—	92,763
Proceeds from sale of business and other assets, net	30,283	6,737
Other	(6,898)	(4,892)
Net cash used in investing activities	$ (59,544)	$ (624,867)
FINANCING ACTIVITIES:		
Payments on borrowings, net	$ (78,745)	$ (96,683)
Other	(26,736)	(11,884)
Net cash used in financing activities	$ (105,481)	$ (108,567)
Effect of foreign exchange rate	285	654
NET INCREASE (DECREASE) IN CASH, CASH EQUIVALENTS, RESTRICTED CASH AND RESTRICTED CASH EQUIVALENTS	$ 246,310	$ (335,388)
CASH, CASH EQUIVALENTS, RESTRICTED CASH AND RESTRICTED CASH EQUIVALENTS, BEGINNING OF PERIOD	1,385,000	1,720,388
CASH, CASH EQUIVALENTS, RESTRICTED CASH AND RESTRICTED CASH EQUIVALENTS, END OF PERIOD	$ 1,631,310	$ 1,385,000

SUPPLEMENTAL FINANCIAL INFORMATION

To supplement the financial measures prepared in accordance with U.S. generally accepted accounting principles (GAAP), the Company uses certain non-GAAP financial measures. For additional information on the Company's use of such non-GAAP financial measures, refer to Endo's Current Report on Form 8-K furnished today to the U.S. Securities and Exchange Commission, which includes an explanation of the Company's reasons for using non-GAAP measures.

The tables below provide reconciliations of certain of the Company's non-GAAP financial measures to their most directly comparable GAAP amounts. Refer to the "Notes to the Reconciliations of GAAP and Non-GAAP Financial Measures" section below for additional details regarding the adjustments to the non-GAAP financial measures detailed throughout this Supplemental Financial Information section.

Reconciliation of EBITDA and Adjusted EBITDA (non-GAAP)

The following table provides a reconciliation of Net (loss) income (GAAP) to Adjusted EBITDA (non-GAAP) for the three months and years ended December 31, 2021 and 2020 (in thousands):

	Three Months Ended December 31,		Year Ended December 31,	
	2021	2020	2021	2020
Net (loss) income (GAAP)	$ (562,062)	$ 119,343	$ (613,245)	$ 183,944
Income tax expense (benefit)	9,106	(149,466)	22,478	(273,982)
Interest expense, net	143,501	135,250	562,353	532,939
Depreciation and amortization (14)	104,254	119,562	432,380	496,349
EBITDA (non-GAAP)	$ (305,201)	$ 224,689	$ 403,966	$ 939,250
Upfront and milestone-related payments (2)	$ 20,245	$ 32,606	$ 26,451	$ 35,075
Amounts related to continuity and separation benefits, cost reductions and strategic review initiatives (3)	32,280	25,926	90,912	126,282
Certain litigation-related and other contingencies, net (4)	226,168	4,889	345,495	(19,049)
Certain legal costs (5)	53,187	15,935	136,148	67,819
Asset impairment charges (6)	364,584	14,147	414,977	120,344
Acquisition-related and integration costs (7)	—	196	414	196
Fair value of contingent consideration (8)	(2,022)	(747)	(8,793)	16,353
Loss on extinguishment of debt (9)	—	—	13,753	—
Share-based compensation (14)	6,990	7,905	29,227	36,167
Other (income) expense, net (15)	(15,103)	4,208	(19,774)	(21,110)
Other (10)	1	(23)	3,882	31,095
Discontinued operations, net of tax (12)	5,395	21,904	44,164	63,520
Adjusted EBITDA (non-GAAP)	$ 386,524	$ 351,635	$ 1,480,822	$ 1,395,942

Reconciliation of Adjusted Income from Continuing Operations (non-GAAP)

The following table provides a reconciliation of the Company's (Loss) income from continuing operations (GAAP) to Adjusted income from continuing operations (non-GAAP) for the three months and years ended December 31, 2021 and 2020 (in thousands):

	Three Months Ended December 31,		Year Ended December 31,	
	2021	2020	2021	2020
(Loss) income from continuing operations (GAAP)	$ (556,667)	$ 141,247	$ (569,081)	$ 247,464
Non-GAAP adjustments:				
Amortization of intangible assets (1)	91,806	101,742	372,907	427,543
Upfront and milestone-related payments (2)	20,245	32,606	26,451	35,075
Amounts related to continuity and separation benefits, cost reductions and strategic review initiatives (3)	32,280	25,926	90,912	126,282
Certain litigation-related and other contingencies, net (4)	226,168	4,889	345,495	(19,049)
Certain legal costs (5)	53,187	15,935	136,148	67,819
Asset impairment charges (6)	364,584	14,147	414,977	120,344
Acquisition-related and integration costs (7)	—	196	414	196
Fair value of contingent consideration (8)	(2,022)	(747)	(8,793)	16,353
Loss on extinguishment of debt (9)	—	—	13,753	—
Other (10)	(15,325)	3,727	(15,870)	17,164
Tax adjustments (11)	(14,222)	(163,673)	(90,964)	(368,821)
Adjusted income from continuing operations (non-GAAP)	$ 200,034	$ 175,995	$ 716,349	$ 670,370

Reconciliation of Other Adjusted Income Statement Data (non-GAAP)

The following tables provide detailed reconciliations of various other income statement data between the GAAP and non-GAAP amounts for the three months and years ended December 31, 2021 and 2020 (in thousands, except per share data):

Three Months Ended December 31, 2021

	Total revenues, net	Cost of revenues	Gross margin	Gross margin %	Total operating expenses	Operating expense to revenue %	Operating (loss) income from continuing operations	Operating margin %	Other non-operating expense, net	(Loss) income from continuing operations before	Income tax expense	Effective tax rate	(Loss) income from continuing operations	Discontinued operations, net of tax	Net (loss) income	Diluted net (loss) income per share from continuing operations
Reported (GAAP)	$ 789,429	$ 311,223	$ 478,206	60.6 %	$ 897,369	113.7 %	$ (419,163)	(53.1)%	$ 128,398	$ (547,561)	$ 9,106	(1.7)%	$ (556,667)	$ (5,395)	$ (562,062)	$ (2.38)
Items impacting comparability:																
Amortization of intangible assets (1)	—	(91,806)	91,806		—		91,806		—	91,806	—		91,806	—	91,806	
Upfront and milestone-related payments (2)	—	(125)	125		(20,120)		20,245		—	20,245	—		20,245	—	20,245	
Amounts related to continuity and separation benefits, cost reductions and strategic review initiatives (3)	—	949	(949)		(33,229)		32,280		—	32,280	—		32,280	—	32,280	
Certain litigation-related and other contingencies, net (4)	—	—	—		(226,168)		226,168		—	226,168	—		226,168	—	226,168	
Certain legal costs (5)	—	—	—		(53,187)		53,187		—	53,187	—		53,187	—	53,187	
Asset impairment charges (6)	—	—	—		(364,584)		364,584		—	364,584	—		364,584	—	364,584	
Fair value of contingent consideration (8)	—	—	—		2,022		(2,022)		—	(2,022)	—		(2,022)	—	(2,022)	
Other (10)	—	—	—		—		—		15,325	(15,325)	—		(15,325)	—	(15,325)	
Tax adjustments (11)	—	—	—		—		—		—	—	14,222		(14,222)	—	(14,222)	
Exclude discontinued operations, net of tax (12)	—	—	—		—		—		—	—	—		—	5,395	5,395	
After considering items (non-GAAP)	$ 789,429	$ 220,241	$ 569,188	72.1 %	$ 202,103	25.6 %	$ 367,085	46.5 %	$ 143,723	$ 223,362	$ 23,328	10.4 %	$ 200,034	$ —	$ 200,034	$ 0.84

Three Months Ended December 31, 2020

	Total revenues, net	Cost of revenues	Gross margin	Gross margin %	Total operating expenses	Operating expense to revenue %	Operating income from continuing operations	Operating margin %	Other non-operating expense, net	(Loss) income from continuing operations before	Income tax (benefit) expense	Effective tax rate	Income from continuing operations	Discontinued operations, net of tax	Net income	Diluted net income per share from continuing operations (13)
Reported (GAAP)	$ 760,221	$ 369,539	$ 390,682	51.4 %	$ 259,443	34.1 %	$ 131,239	17.3 %	$ 139,458	$ (8,219)	$ (149,466)	1,818.5 %	$ 141,247	$ (21,904)	$ 119,343	$ 0.60
Items impacting comparability:																
Amortization of intangible assets (1)	—	(101,742)	101,742		—		101,742		—	101,742	—		101,742	—	101,742	
Upfront and milestone-related payments (2)	—	(925)	925		(31,681)		32,606		—	32,606	—		32,606	—	32,606	
Amounts related to continuity and separation benefits, cost reductions and strategic review initiatives (3)	—	(11,721)	11,721		(14,205)		25,926		—	25,926	—		25,926	—	25,926	
Certain litigation-related and other contingencies, net (4)	—	—	—		(4,889)		4,889		—	4,889	—		4,889	—	4,889	
Certain legal costs (5)	—	—	—		(15,935)		15,935		—	15,935	—		15,935	—	15,935	
Asset impairment charges (6)	—	—	—		(14,147)		14,147		—	14,147	—		14,147	—	14,147	
Acquisition-related and integration costs (7)	—	—	—		(196)		196		—	196	—		196	—	196	
Fair value of contingent consideration (8)	—	—	—		747		(747)		—	(747)	—		(747)	—	(747)	
Other (10)	—	—	—		—		—		(3,727)	3,727	—		3,727	—	3,727	
Tax adjustments (11)	—	—	—		—		—		—	—	163,673		(163,673)	—	(163,673)	
Exclude discontinued operations, net of tax (12)	—	—	—		—		—		—	—	—		—	21,904	21,904	
After considering items (non-GAAP)	$ 760,221	$ 255,151	$ 505,070	66.4 %	$ 179,137	23.6 %	$ 325,933	42.9 %	$ 135,731	$ 190,202	$ 14,207	7.5 %	$ 175,995	$ —	$ 175,995	$ 0.75

	Year Ended December 31, 2021															
	Total revenues, net	Cost of revenues	Gross margin	Gross margin %	Total operating expenses	Operating expense to revenue %	Operating income from continuing operations	Operating margin %	Other non-operating expense, net	(Loss) income from continuing operations before	Income tax expense	Effective tax rate	(Loss) income from continuing operations	Discontinued operations, net of tax	Net (loss) income	Diluted net (loss) income per share from continuing operations
Reported (GAAP)	$2,993,206	$1,221,064	$1,772,142	59.2 %	$1,762,413	58.9 %	$ 9,729	0.3 %	$ 556,332	$ (546,603)	$ 22,478	(4.1)%	$ (569,081)	$ (44,164)	$ (613,245)	$ (2.44)
Items impacting comparability:																
Amortization of intangible assets (1)	—	(372,907)	372,907		—		372,907		—	372,907	—		372,907	—	372,907	
Upfront and milestone-related payments (2)	—	(1,301)	1,301		(25,150)		26,451		—	26,451	—		26,451	—	26,451	
Amounts related to continuity and separation benefits, cost reductions and strategic review initiatives (3)	—	(9,058)	9,058		(81,854)		90,912		—	90,912	—		90,912	—	90,912	
Certain litigation-related and other contingencies, net (4)	—	—	—		(345,495)		345,495		—	345,495	—		345,495	—	345,495	
Certain legal costs (5)	—	—	—		(136,148)		136,148		—	136,148	—		136,148	—	136,148	
Asset impairment charges (6)	—	—	—		(414,977)		414,977		—	414,977	—		414,977	—	414,977	
Acquisition-related and integration costs (7)	—	—	—		(414)		414		—	414	—		414	—	414	
Fair value of contingent consideration (8)	—	—	—		8,793		(8,793)		—	(8,793)	—		(8,793)	—	(8,793)	
Loss on extinguishment of debt (9)	—	—	—		—		—		(13,753)	13,753	—		13,753	—	13,753	
Other (10)	—	—	—		(3,879)		3,879		19,749	(15,870)	—		(15,870)	—	(15,870)	
Tax adjustments (11)	—	—	—		—		—		—	—	90,964		(90,964)	—	(90,964)	
Exclude discontinued operations, net of tax (12)	—	—	—		—		—		—	—	—		—	44,164	44,164	
After considering items (non-GAAP)	$2,993,206	$ 837,798	$2,155,408	72.0 %	$ 763,289	25.5 %	$1,392,119	46.5 %	$ 562,328	$ 829,791	$ 113,442	13.7 %	$ 716,349	$ —	$ 716,349	$ 3.03

	Year Ended December 31, 2020															
	Total revenues, net	Cost of revenues	Gross margin	Gross margin %	Total operating expenses	Operating expense to revenue %	Operating income from continuing operations	Operating margin %	Other non-operating expense, net	(Loss) income from continuing operations before	Income tax (benefit) expense	Effective tax rate	Income from continuing operations	Discontinued operations, net of tax	Net income	Diluted net income per share from continuing operations (13)
Reported (GAAP)	$2,903,074	$1,442,511	$1,460,563	50.3 %	$ 975,252	33.6 %	$ 485,311	16.7 %	$ 511,829	$ (26,518)	$ (273,982)	1,033.2 %	$ 247,464	$ (63,520)	$ 183,944	$ 1.06
Items impacting comparability:																
Amortization of intangible assets (1)	—	(427,543)	427,543		—		427,543		—	427,543	—		427,543	—	427,543	
Upfront and milestone-related payments (2)	—	(1,717)	1,717		(33,358)		35,075		—	35,075	—		35,075	—	35,075	
Amounts related to continuity and separation benefits, cost reductions and strategic review initiatives (3)	—	(55,413)	55,413		(70,869)		126,282		—	126,282	—		126,282	—	126,282	
Certain litigation-related and other contingencies, net (4)	—	—	—		19,049		(19,049)		—	(19,049)	—		(19,049)	—	(19,049)	
Certain legal costs (5)	—	—	—		(67,819)		67,819		—	67,819	—		67,819	—	67,819	
Asset impairment charges (6)	—	—	—		(120,344)		120,344		—	120,344	—		120,344	—	120,344	
Acquisition-related and integration costs (7)	—	—	—		(196)		196		—	196	—		196	—	196	
Fair value of contingent consideration (8)	—	—	—		(16,353)		16,353		—	16,353	—		16,353	—	16,353	
Other (10)	—	—	—		(31,118)		31,118		13,954	17,164	—		17,164	—	17,164	
Tax adjustments (11)	—	—	—		—		—		—	—	368,821		(368,821)	—	(368,821)	
Exclude discontinued operations, net of tax (12)	—	—	—		—		—		—	—	—		—	63,520	63,520	
After considering items (non-GAAP)	$2,903,074	$ 957,838	$1,945,236	67.0 %	$ 654,244	22.5 %	$1,290,992	44.5 %	$ 525,783	$ 765,209	$ 94,839	12.4 %	$ 670,370	$ —	$ 670,370	$ 2.87

Notes to the Reconciliations of GAAP and Non-GAAP Financial Measures

Notes to certain line items included in the reconciliations of the GAAP financial measures to the non-GAAP financial measures for the three months and years ended December 31, 2021 and 2020 are as follows:

(1) To exclude amortization expense related to intangible assets.

(2) Adjustments for upfront and milestone-related payments to partners included the following (in thousands):

	Three Months Ended December 31,			
	2021		2020	
	Cost of revenues	Operating expenses	Cost of revenues	Operating expenses
Sales-based	$ 125	$ —	$ 925	$ —
Development-based	—	20,120	—	31,681
Total	$ 125	$ 20,120	$ 925	$ 31,681

	Year Ended December 31,			
	2021		2020	
	Cost of revenues	Operating expenses	Cost of revenues	Operating expenses
Sales-based	$ 1,301	$ —	$ 1,717	$ —
Development-based	—	25,150	—	33,358
Total	$ 1,301	$ 25,150	$ 1,717	$ 33,358

(3) Adjustments for amounts related to continuity and separation benefits, cost reductions and strategic review initiatives included the following (in thousands):

	Three Months Ended December 31,			
	2021		2020	
	Cost of revenues	Operating expenses	Cost of revenues	Operating expenses
Continuity and separation benefits	$ (3,119)	$ 13,100	$ 3,585	$ 7,451
Accelerated depreciation	1,715	672	5,039	2,744
Other, including strategic review initiatives	455	19,457	3,097	4,010
Total	$ (949)	$ 33,229	$ 11,721	$ 14,205

	Year Ended December 31,			
	2021		2020	
	Cost of revenues	Operating expenses	Cost of revenues	Operating expenses
Continuity and separation benefits	$ (16,946)	$ 25,760	$ 36,775	$ 50,132
Accelerated depreciation	19,037	5,680	15,567	6,892
Other, including strategic review initiatives	6,967	50,414	3,071	13,845
Total	$ 9,058	$ 81,854	$ 55,413	$ 70,869

The amounts in the tables above include adjustments related to previously announced restructuring activities, certain continuity and transitional compensation arrangements, certain other cost reduction initiatives and certain strategic review initiatives.

(4) To exclude adjustments to accruals for litigation-related settlement charges and certain settlement proceeds related to suits filed by subsidiaries.

(5) To exclude opioid-related legal expenses.

(6) Adjustments for asset impairment charges included the following (in thousands):

	Three Months Ended December 31,		Year Ended December 31,	
	2021	2020	2021	2020
Goodwill impairment charges	$ 363,000	$ —	$ 363,000	$ 32,786
Other intangible asset impairment charges	—	14,146	7,811	79,917
Property, plant and equipment impairment charges	1,584	1	2,011	1,249
Operating lease right-of-use asset impairment charges	—	—	—	6,392
Disposal group impairment charges	—	—	42,155	—
Total	$ 364,584	$ 14,147	$ 414,977	$ 120,344

(7) To exclude integration costs.

(8) To exclude the impact of changes in the fair value of contingent consideration liabilities resulting from changes to estimates regarding the timing and amount of the future revenues of the underlying products and changes in other assumptions impacting the probability of incurring, and extent to which the Company could incur, related contingent obligations.

(9) To exclude the loss on the extinguishment of debt associated with the Company's March 2021 refinancing transactions.

(10) The "Other" rows included in each of the above reconciliations of GAAP financial measures to non-GAAP financial measures (except for the reconciliations of Net (loss) income (GAAP) to Adjusted EBITDA (non-GAAP)) include the following (in thousands):

	Three Months Ended December 31,			
	2021		2020	
	Operating expenses	Other non-operating expenses	Operating expenses	Other non-operating expenses
Foreign currency impact related to the re-measurement of intercompany debt instruments	$ —	$ 331	$ —	$ 4,345
Gain on sale of business and other assets	—	(5,085)	—	—
Other miscellaneous	—	(10,571)	—	(618)
Total	$ —	$ (15,325)	$ —	$ 3,727

	Year Ended December 31,			
	2021		2020	
	Operating expenses	Other non-operating expenses	Operating expenses	Other non-operating expenses
Foreign currency impact related to the re-measurement of intercompany debt instruments	$ —	$ 797	$ —	$ 1,919
Gain on sale of business and other assets	—	(5,085)	—	(11,325)
Debt modification costs	3,879	—	31,118	—
Other miscellaneous	—	(15,461)	—	(4,548)
Total	$ 3,879	$ (19,749)	$ 31,118	$ (13,954)

The 2021 amounts in the "Other miscellaneous" rows of the tables above primarily relate to gains associated with the termination of certain contracts.

The "Other" row included in the reconciliations of Net (loss) income (GAAP) to Adjusted EBITDA (non-GAAP) primarily relates to the items enumerated in the foregoing "Operating expenses" columns.

(11) Adjusted income taxes are calculated by tax effecting adjusted pre-tax income and permanent book-tax differences at the applicable effective tax rate that will be determined by reference to statutory tax rates in the relevant jurisdictions in which the Company operates. Adjusted income taxes include current and deferred income tax expense commensurate with the non-GAAP measure of profitability.

(12) To exclude the results of the businesses reported as discontinued operations, net of tax.

(13) Calculated as income or loss from continuing operations divided by the applicable weighted average share number. The applicable weighted average share numbers are as follows (in thousands):

	Three Months Ended December 31,		Year Ended December 31,	
	2021	2020	2021	2020
GAAP	233,681	234,474	232,785	233,653
Non-GAAP Adjusted	237,045	234,474	236,665	233,653

(14) Depreciation and amortization and Share-based compensation per the Adjusted EBITDA reconciliations do not include amounts reflected in other lines of the reconciliations, including Amounts related to continuity and separation benefits, cost reductions and strategic review initiatives.

(15) To exclude Other (income) expense, net per the Consolidated Statements of Operations.

Reconciliation of Net Debt Leverage Ratio (non-GAAP)

The following table provides a reconciliation of the Company's Net loss (GAAP) to Adjusted EBITDA (non-GAAP) for the twelve months ended December 31, 2021 (in thousands) and the calculation of the Company's Net Debt Leverage Ratio (non-GAAP):

	Twelve Months Ended December 31, 2021
Net loss (GAAP)	$ (613,245)
Income tax expense	22,478
Interest expense, net	562,353
Depreciation and amortization (14)	432,380
EBITDA (non-GAAP)	$ 403,966
Upfront and milestone-related payments	$ 26,451
Amounts related to continuity and separation benefits, cost reductions and strategic review initiatives	90,912
Certain litigation-related and other contingencies, net	345,495
Certain legal costs	136,148
Asset impairment charges	414,977
Acquisition-related and integration costs	414
Fair value of contingent consideration	(8,793)
Loss on extinguishment of debt	13,753
Share-based compensation (14)	29,227
Other income, net	(19,774)
Other	3,882
Discontinued operations, net of tax	44,164
Adjusted EBITDA (non-GAAP)	$ 1,480,822
Calculation of Net Debt:	
Debt	$ 8,249,322
Cash (excluding Restricted Cash)	1,507,196
Net Debt (non-GAAP)	$ 6,742,126
Calculation of Net Debt Leverage:	
Net Debt Leverage Ratio (non-GAAP)	4.6

Non-GAAP Financial Measures

The Company utilizes certain financial measures that are not prescribed by or prepared in accordance with accounting principles generally accepted in the U.S. (GAAP). These non-GAAP financial measures are not, and should not be viewed as, substitutes for GAAP net income and its components and diluted net income per share amounts. Despite the importance of these measures to management in goal setting and performance measurement, the company stresses that these are non-GAAP financial measures that have no standardized meaning prescribed by GAAP and, therefore, have limits in their usefulness to investors. Because of the non-standardized definitions, non-GAAP adjusted EBITDA and non-GAAP adjusted net income from continuing operations and its components (unlike GAAP net income from continuing operations and its components) may not be comparable to the calculation of similar measures of other companies. These non-GAAP financial measures are presented solely to permit investors to more fully understand how management assesses performance.

Investors are encouraged to review the reconciliations of the non-GAAP financial measures used in this press release to their most directly comparable GAAP financial measures. However, the Company does not provide reconciliations of projected non-GAAP financial measures to GAAP financial measures, nor does it provide comparable projected GAAP financial measures for such projected non-GAAP financial measures. The Company is unable to provide such reconciliations without unreasonable efforts due to the inherent difficulty in forecasting and quantifying certain amounts that are necessary for such reconciliations, including adjustments that could be made for asset impairments, contingent consideration adjustments, legal settlements, gain / loss on extinguishment of debt, adjustments to inventory and other charges reflected in the reconciliation of historic numbers, the amounts of which could be significant.

See Endo's Current Report on Form 8-K furnished today to the U.S. Securities and Exchange Commission for an explanation of Endo's non-GAAP financial measures.

About Endo International plc

Endo (NASDAQ: ENDP) is a specialty pharmaceutical company committed to helping everyone we serve live their best life through the delivery of quality, life-enhancing therapies. Our decades of proven success come from a global team of passionate employees collaborating to bring the best treatments forward. Together, we boldly transform insights into treatments benefiting those who need them, when they need them. Learn more at www.endo.com or connect with us on LinkedIn.

Cautionary Note Regarding Forward-Looking Statements

Certain information in this press release may be considered "forward-looking statements" within the meaning of the Private Securities Litigation Reform Act of 1995 and any applicable Canadian securities legislation, including, but not limited to, the statements by Mr. Coleman, as well as other statements regarding product development, product launches, product demand and market potential; the expansion and enhancement of our product portfolio; progress on our strategic priorities; the status and outcome of litigation; financial guidance or outlook for the first quarter of 2022, full-year 2022 or any other future periods; the impact of and response to the COVID-19 pandemic; the status of our contingency planning and strategic review, including any potential restructuring or bankruptcy filing; and any other statements that refer to our expected, estimated or anticipated future results or that do not relate solely to historical facts. Statements including words or phrases such as "believe," "expect," "anticipate," "intend," "estimate," "plan," "will," "may," "look forward," "intend," "guidance," "future," "potential" or similar expressions are forward-looking statements. Because forecasts are inherently estimates that cannot be made with precision, Endo's performance at times differs materially from its estimates and targets, and Endo often does not know what the actual results will be until after the end of the applicable reporting period. Therefore, Endo will not report or comment on its progress during a current quarter except through public announcement. Any statement made by others with respect to progress during a current quarter cannot be attributed to Endo. All forward-looking statements in this press release reflect Endo's current analysis of existing trends and information and represent Endo's judgment only as of the date of this press release. Actual results may differ materially and adversely from current expectations based on a number of factors affecting Endo's businesses, including, among other things, the following: the outcome of our strategic review, contingency planning and any potential restructuring or bankruptcy filing; the timing, impact or results of any pending or future litigation, investigations, proceedings or claims, including opioid-related matters and tax-related matters; actual or contingent liabilities; settlement discussions or negotiations; the impact of competition, including the loss of exclusivity and generic competition for VASOSTRICT®; our ability to satisfy judgments or settlements or pursue appeals including bonding requirements; our ability to adjust to changing market conditions; our ability to attract and retain key personnel; our inability to maintain compliance with financial covenants and operating obligations which would expose us to potential events of default under our outstanding indebtedness; our ability to incur additional debt or equity financing for working capital, capital expenditures, business development, debt service requirements, acquisitions or general corporate or other purposes; our ability to refinance our indebtedness; a significant reduction in our short-term or long-term revenues which could cause us to be unable to fund our operations and liquidity needs or repay indebtedness; supply chain interruptions or difficulties; changes in competitive or market conditions; changes in legislation or regulatory developments; our ability to obtain and maintain adequate protection for our intellectual property rights; the timing and uncertainty of the results of both the research and development and regulatory processes, including regulatory decisions, product recalls, withdrawals and other unusual items; domestic and foreign health care and cost containment reforms, including government pricing, tax and reimbursement policies; technological advances and patents obtained by competitors; the performance, including the approval, introduction, and consumer and physician acceptance of new products and the continuing acceptance of currently marketed products; the impact that known and unknown side effects may have on market perception and consumer preference for our products; the effectiveness of advertising and other

promotional campaigns; the timely and successful implementation of any strategic initiatives; unfavorable publicity regarding the misuse of opioids; the uncertainty associated with the identification of and successful consummation and execution of external corporate development initiatives and strategic partnering transactions; our ability to advance our strategic priorities, develop our product pipeline and continue to develop the market for QWO® and other products; and our ability to obtain and successfully manufacture, maintain and distribute a sufficient supply of products to meet market demand in a timely manner. In addition, U.S. and international economic conditions, including consumer confidence and debt levels, taxation, changes in interest and currency exchange rates, international relations, capital and credit availability, the status of financial markets and institutions, the impact of and response to the ongoing COVID-19 pandemic and the impact of continued economic volatility, can materially affect our results. The occurrence or possibility of any such result has caused us to engage, and may result in further engagement in strategic reviews that ultimately may result in our pursuing one or more significant corporate transactions or other remedial measures, including on a preventative or proactive basis. Those remedial measures could include a potential bankruptcy filing (which, if it occurred, would subject us to additional risks and uncertainties that could adversely affect our business prospects and ability to continue as a going concern), corporate reorganization or restructuring activities involving all or a portion of our business, asset sales or other divestitures, cost-saving initiatives or other corporate realignments, seeking strategic partnerships and exiting certain product or geographic markets. Some of these measures could take significant time to implement and others may require judicial or other third-party approval. Any such actions may be complex, could entail significant costs and charges or could otherwise negatively impact shareholder value, and there can be no assurance that we will be able to accomplish any of these alternatives on terms acceptable to us, or at all, or that they will result in their intended benefits. Therefore, the reader is cautioned not to rely on these forward-looking statements. Endo expressly disclaims any intent or obligation to update these forward-looking statements, except as required to do so by law.

Additional information concerning risk factors, including those referenced above, can be found in press releases issued by Endo, as well as Endo's public periodic filings with the U.S. Securities and Exchange Commission and with securities regulators in Canada, including the discussion under the heading "Risk Factors" in Endo's most recent Annual Report on Form 10-K and any subsequent Quarterly Reports on Form 10-Q or other filings with the U.S. Securities and Exchange Commission. Copies of Endo's press releases and additional information about Endo are available at www.endo.com or you can contact the Endo Investor Relations Department by calling 845-364-4833.

SOURCE Endo International plc

Media: Heather Zoumas-Lubeski, (484) 216-6829; Investors: Pravesh Khandelwal, (845) 364-4833

#####

Evidence review: Gender-affirming hormones for children and adolescents with gender dysphoria

This document will help inform Dr Hilary Cass' independent review into gender identity services for children and young people. It was commissioned by NHS England and Improvement who commissioned the Cass review. It aims to assess the evidence for the clinical effectiveness, safety and cost-effectiveness of gender-affirming hormones for children and adolescents aged 18 years or under with gender dysphoria.

The document was prepared by NICE in October 2020.

The content of this evidence review was up to date on 21 October 2020. See summaries of product characteristics (SPCs), British National Formulary (BNF) or the Medicines and Healthcare products Regulatory Agency (MHRA) or NICE websites for up-to-date information.

Contents

1. Introduction

This review aims to assess the evidence for the clinical effectiveness, safety and cost-effectiveness of gender-affirming hormones for children and adolescents aged 18 years or under with gender dysphoria. The review follows the NHS England Specialised Commissioning process and template and is based on the criteria outlined in the PICO framework (see appendix A). This document will help inform Dr Hilary Cass' independent review into gender identity services for children and young people.

Gender dysphoria in children, also known as gender identity disorder or gender incongruence of childhood (World Health Organisation 2020), refers to discomfort or distress that is caused by a discrepancy between a person's gender identity (how they see themselves[1] regarding their gender) and that person's sex assigned at birth and the associated gender role, and/or primary and secondary sex characteristics (Diagnostic and Statistical Manual of Mental Disorders 2013).

Gender-affirming hormones are oestradiol for sex assigned at birth males (transfemales) and testosterone for sex assigned at birth females (transmales). The aim of gender-affirming hormones is to induce the development of the physical sex characteristics congruent with the individual's gender expression while aiming to improve mental health and quality of life outcomes.

No oestradiol-containing products are licensed for gender dysphoria and therefore any use for children and adolescents with gender dysphoria is off-label.

The only testosterone-containing product licensed for gender dysphoria is Sustanon 250 mg/ml solution for injection, which is indicated as supportive therapy for transmales, use of all other testosterone-containing products for children and adolescents with gender dysphoria is off-label.

For children and adolescents with gender dysphoria it is recommended that management plans are tailored to the needs of the individual and aim to ameliorate the potentially negative impact of gender dysphoria on general developmental processes, to support young people and their families in managing the uncertainties inherent in gender identity development and to provide ongoing opportunities for exploration of gender identity. The plans may also include psychological support and exploration and, for some individuals, the use of gonadotrophin releasing hormone (GnRH) analogues in adolescence to suppress puberty; this may be followed later with gender-affirming hormones of the desired sex (NHS England 2013).

Currently NHS England, as part of the Gender Identity Development Service for Children and Adolescents, routinely commissions gender-affirming hormones for young people with continuing gender dysphoria from around their 16th birthday subject to individuals meeting the eligibility and readiness criteria (Clinical Commissioning Policy 2016).

[1] Gender refers to the roles, behaviours, activities, attributes and opportunities that any society considers appropriate for girls and boys, and women and men (World Health Organisation, Health Topics: Gender).

2. Executive summary of the review

Ten observational studies were included in the evidence review. Seven studies were retrospective observational studies (Allen et al. 2019, Kaltiala et al. 2020, Khatchadourian et al. 2014, Klaver et Al. 2020, Klink et al. 2015, Stoffers et al. 2019, Vlot et al. 2017) and 3 studies were prospective longitudinal observational studies (Achille et al. 2020, Kuper et al. 2020, Lopez de Lara et al. 2020). No studies directly compared gender-affirming hormones to a control group (either placebo or active comparator). Follow-up was relatively short across all studies, with an average duration of treatment with gender-affirming hormones between around 1 year and 5.8 years.

The terminology used in this topic area is continually evolving and is different depending on stakeholder perspectives. In this evidence review we have used the phrase 'people's assigned sex at birth' rather than saying natal or biological sex and 'cross sex hormones' are now referred to as 'gender-affirming hormones'. The research studies may use historical terms which are no longer considered appropriate.

In children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Critical outcomes

The critical outcomes for decision making are impact on gender dysphoria, impact on mental health and quality of life. The quality of evidence for all these outcomes was assessed as very low certainty using modified GRADE.

Impact on gender dysphoria
The study by Lopez de Lara et al. 2020 in 23 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, gender dysphoria (measured using the Utrecht Gender Dysphoria Scale [UGDS]) was statistically significantly reduced (improved) from a mean [±SD] score of 57.1 (±4.1) points at baseline to 14.7 (±3.2) points at 12 months, which is below the threshold (40 points) for gender dysphoria ($p<0.001$).

Impact on mental health
Depression
The study by Lopez de Lara et al. 2020 in 23 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, depression (measured using the Beck Depression Inventory-II [BDI-II]) was statistically significantly reduced from a mean [±SD] score of 19.3 (±5.5) points at baseline to 9.7 (±3.9) points at 12 months ($p<0.001$).

The study by Achille et al. 2020 in 50 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, depression was statistically significantly reduced from baseline to about 12 months follow-up:

- The Center for Epidemiologic Studies Depression (CESD-R) improved from a mean score of 21.4 points at baseline to 13.9 points ($p<0.001$).
- The Patient Health Questionnaire (PHQ 9) Modified for Teens improved, although absolute scores were not reported numerically ($p<0.001$).

The study by Kuper et al. 2020 in 148 adolescents with gender dysphoria (of whom 123 received gender-affirming hormones) found that during treatment with gender-affirming hormones for an average of 10.9 months, the impact on depression (measured using the Quick Inventory of Depressive Symptoms [QIDS]) was unclear as no statistical analysis was reported. The mean (±SD) self-reported score was 9.6 points (±5.0) at baseline and 7.4 (±4.5) at follow-up. The mean (±SD) clinician-reported score was 5.9 points (±4.1) at baseline and 6.0 (±3.8).

The study by Kaltiala et al. 2020 in 52 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, statistically significantly fewer participants needed treatment for depression (54% at initial assessment compared with 15% at 12-month follow-up, p<0.001). No details of the treatments for depression are reported.

Anxiety

The study by Lopez de Lara et al. 2020 in 23 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, state anxiety (measured using the State-Trait Anxiety Inventory [STAI] – State subscale) was statistically significantly reduced from a mean (±SD) score of 33.3 points (±9.1) at baseline to 16.8 points (±8.1) at 12 months (p<0.001). Trait anxiety (measured using STAI – Trait subscale) was also statistically significantly reduced from a mean (±SD) score of 33.0 (±7.2) points at baseline to 18.5 (±8.4) points at 12 months (p<0.001).

The study by Kuper et al. 2020 in 148 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, small reductions were seen in anxiety, panic, generalised anxiety, social anxiety and separation anxiety symptoms and school avoidance (measured using the Screen for Child Anxiety Related Emotional Disorders [SCARED] questionnaire) from baseline to follow-up (mean duration of treatment 10.9 months). The statistical significance of these findings are unknown as no statistical analyses were reported.

The study by Kaltiala et al. 2020 in 52 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, statistically significantly fewer participants needed treatment for anxiety (48% at initial assessment compared with 15% at 12-month follow-up, p<0.001). No details of treatments for anxiety are reported.

Suicidality and self-injury

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, suicide risk (measured using the Ask Suicide-Screening Questions [ASQ]) was statistically significantly reduced from an adjusted mean (±SE) score of 1.11 points (±0.22) at baseline to 0.27 points (±0.12) after about 12 months (p<0.001).

The study by Achille et al. 2020 in 50 adolescents with gender dysphoria (of whom 35 received gender-affirming hormones at follow-up) found that during treatment with gender-affirming hormones, the impact on suicidal ideation was unclear (measured using the PHQ 9_Modified for Teens with additional questions for suicidal ideation). At baseline 10%of participants had suicidal ideation and 6% had suicidal ideation after about 12 months, but it is unclear if these participants received gender-affirming hormones. No statistical analyses were reported.

The study by Kuper et al. 2020 in 148 adolescents with gender dysphoria reported the impact on suicidal ideation, suicide attempts and non-suicidal self-injury during treatment with gender-affirming hormones, after mean 10.9 months follow-up . The statistical significance of these findings are unknown as no statistical analyses were reported:

- Suicidal ideation was reported in 25% of participants 1 month before the initial assessment and in 38% of participants during follow-up.
- Suicide attempts were reported in 2% of participants at 3 months before the initial assessment and in 5% during follow-up.
- Self-injury was reported in 10% of participants at 3 months before the initial assessment and in 17% during follow-up.

The study by Kaltiala et al. 2020 in 52 adolescents with gender dysphoria reported that during treatment with gender-affirming hormones, statistically significantly fewer participants needed treatment for suicidal ideation or self-harm (35% at initial assessment compared with 4% at 12-month follow-up, p<0.001). No details of treatments for suicidal ideation or self-harm are reported.

Other related symptoms

The study by Kaltiala et al. 2020 in 52 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, there was no statistically significant difference in the number of people needing treatment for either psychotic symptoms or psychosis, conduct problems or antisocial behaviour, substance abuse, autism, attention deficit hyperactivity disorder (ADHD) or eating disorders during the 12-month 'real life' phase compared with before or during the assessment. No details of the treatments received are reported.

Impact on quality of life

The study by Achille et al. 2020 in 50 adolescents with gender dysphoria (of whom 35 were receiving gender-affirming hormones at follow-up) found that during treatment with gender-affirming hormones, quality of life (measured using the Quality of Life Enjoyment and Satisfaction Questionnaire [QLES-Q-SF]) was statistically significantly improved from baseline to about 12 months, but absolute scores were not reported numerically (p<0.001).

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, quality of life (measured using the General Well-Being Scale [GWBS] of the Paediatric Quality of Life Inventory) was statistically significantly improved from an adjusted mean (±SE) score of 61.70 (±2.43) points at baseline to 70.23 (±2.15) points at about 12 months (p<0.002).

Important outcomes

The important outcomes for decision making are impact on body image, psychosocial impact, engagement with healthcare services, impact on extent of and satisfaction with surgery and de-transition. The quality of evidence for all these outcomes was assessed as very low certainty using modified GRADE.

Impact on body image

The study by Kuper et al. 2020 in 148 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, the impact on body image is unclear (measured using the Body Image Scale [BIS]). The mean (±SD) BIS score was 70.7 points (±15.2) at baseline and 51.4 points (±18.3) at follow-up (mean duration of treatment 10.9 months; no statistical analysis was reported).

Psychosocial impact
The study by Lopez de Lara et al. 2020 in 23 adolescents with gender dysphoria found that during treatment with gender affirming hormones, family functioning is unchanged (measured using the Family Adaptability, Partnership, Growth, Affection and Resolve [APGAR] test). The mean score was 17.9 points at baseline and 18.0 points at 12-month follow-up (no statistical analysis was reported).

The study by Lopez de Lara et al. 2020 in 23 adolescents with gender dysphoria found that during treatment with gender affirming hormones, behavioural problems (measured using the Strengths and Difficulties Questionnaire [SDQ]) were statistically significantly improved from a mean (±SD) of 14.7 (±3.3) points at baseline to 10.3 points (±2.9) at 12-month follow-up ($p<0.001$).

The study by Kaltiala et al. 2020 in 52 adolescents with gender dysphoria found that about 12-months after starting treatment with gender-affirming hormones:

- Statistically significantly fewer participants were living with parents or guardians (73% versus 40%, $p=0.001$) and statistically significantly fewer participants had normal peer contacts (89% versus 81%, $p<0.001$).
- There were no statistically significant differences in:
    - progress in school or work (64% versus 60%, $p=0.69$),
    - the number of participants who had been dating or in steady relationships (62% versus 58%, $p=0.51$)
    - the ability to cope with matters outside of the home (for example, shopping and travelling alone on local public transport; 81% versus 81%, $p=1.0$)

Engagement with health care services
No evidence was identified.

Impact on extent of and satisfaction with surgery
No evidence was identified.

De-transition
No evidence was identified.

In children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Important outcomes

The important outcomes for decision making are short- and long-term safety outcomes and adverse effects. The quality of evidence for all these outcomes was assessed as very low certainty using modified GRADE.

Bone density

The study by Klink et al. 2015 in 34 adolescents with gender dysphoria (who were previously treated with a GnRH analogue) found that gender-affirming hormones may increase lumbar spine and femoral neck bone density. However, not all results are statistically significant (particularly in transfemales). Z-scores suggest the average bone density at the end of follow-up was generally lower than in the equivalent cisgender population (transfemales compared with cis-males and transmales compared with cis-females). From starting gender-affirming hormones to age 22 years:

- There was no statistically significant difference in lumbar spine bone mineral apparent density (BMAD) z-score in transfemales, but this was statistically significantly higher in transmales (z-score [±SD]: start of hormones -0.50 [±0.81], age 22 years -0.033 [±0.95], p=0.002).
- There was no statistically significant difference in lumbar spine bone mineral density (BMD) z-score in transfemales or transmales.
- Actual lumbar spine BMAD and BMD values were statistically significantly higher in transfemales and transmales.
- There was no statistically significant difference in femoral neck BMD z-score in transfemales, but this was statistically significantly higher in transmales (z-score [SD]: start of hormones -0.35 [0.79], age 22 years -0.35 [0.74], p=0.006).
- There was no statistically significant difference in actual femoral neck BMAD values in transfemales, but this was statistically significantly higher in transmales.
- Actual femoral neck BMD values were statistically significantly higher in transfemales and transmales.

The study by Vlot et al. 2017 in 70 adolescents with gender dysphoria (who were previously treated with a GnRH analogue) found that gender-affirming hormones may increase lumbar spine and femoral neck bone density. However, not all results are statistically significant. Z-scores suggest the average bone density at the end of follow-up was generally lower than the equivalent cisgender population (transfemales compared with cis-males and transmales compared with cis-females). From starting gender-affirming hormones to 24-month follow-up:

- The z-score for lumbar spine BMAD was statistically significantly higher in transfemales with a bone age of less than 15 years (z-score [range]: start of hormones -1.52 [-2.36 to 0.42], 24-month follow-up -1.10 [-2.44 to 0.69], p≤ 0.05) and 15 years and older (z-score [range]: start of hormones -1.15 [-2.21 to 0.08], 24-month follow-up -0.66 [-1.66 to 0.54], p≤ 0.05).
- The z-score for lumbar spine BMAD was statistically significantly higher in transmales with a bone age of less than 14 years (z-score [range]: start of hormones -0.84 [-2.2 to 0.87], 24-month follow-up -0.15 [-1.38 to 0.94], p≤ 0.01) and 14 years and older (z-score [range]: start of hormones -0.29 [-2.28 to 0.90], 24-month follow-up -0.06 [-1.75 to 1.61], p≤ 0.01).
- Actual lumbar spine BMAD values were statistically significantly higher in transfemales and transmales of all bone ages.
- There was no statistically significant difference in femoral neck BMAD z-score in transfemales (all bone ages).

- The z-score for femoral neck BMAD was statistically significantly higher in transmales with a bone age of less than 14 years (z-score [range]: start of hormones -0.37 [-2.28 to 0.47], 24-month follow-up -0.37 [-2.03 to 0.85], p≤ 0.01) and 14 years and older (z-score [range]: start of hormones -0.27 [-1.91 to 1.29], 24-month follow-up 0.02 [-2.1 to 1.35], p≤0.05).
- There was no statistically significant difference in actual femoral neck BMAD values in transfemales (all bone ages), but this was statistically significantly higher in transmales (all bone ages).

The study by Stoffers et al. 2019 in 62 sex assigned at birth females (transmales) with gender dysphoria (who were previously treated with a GnRH analogue) found that during treatment with gender-affirming hormones there was no statistically significant difference in lumbar spine or femoral neck bone density (measured as BMD z-scores or actual values) from starting gender-affirming hormones to any timepoint (6, 12 and 24 months).

Change in clinical parameters

The study by Klaver et al. 2020 in 192 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, from starting treatment to age 22 years:

- Glucose levels, insulin levels and insulin resistance were largely unchanged in transfemales and transmales.
- Total cholesterol, HDL cholesterol and LDL cholesterol levels were unchanged in transfemales, and there was a statistically significant improvement in triglyceride levels.
- Total cholesterol, HDL cholesterol, LDL cholesterol and triglyceride levels significantly worsened in transmales, but mean levels were within the UK reference range at the end of treatment.
- Diastolic blood pressure was statistically significantly increased in transfemales and transmales. Systolic blood pressure was also statistically significantly increased in transmales, but not in transfemales. The absolute increases in blood pressure were small.
- Body mass index was statistically significantly increased in transfemales and transmales, although most participants were within the healthy weight range (18.5 to 24.9 kg/m).

The study by Stoffers et al. 2019 in 62 sex assigned at birth females (transmales) with gender dysphoria found that during treatment with gender affirming hormones, from starting treatment to 24-month follow-up:

- There was no statistically significant change in glycosylated haemoglobin (HbA1c).
- There was no statistically significant change in aspartate aminotransferase (AST), alanine aminotransferase (ALT) and gamma-glutamyltransferase (GCT).
- There was a statistically significant increase in alkaline phosphatase (ALP) at some timepoints, but the difference was not statistically significant by 24-months.
- There was a statistically significant increase in serum creatinine levels at all timepoints up to 24 months, but these were within the UK reference range. Serum urea levels were unchanged (follow-up duration not reported).

Treatment discontinuation and adverse effects

The study by Khatchadourian et al. 2014 in 63 adolescents (24 transfemales and 39 transmales) with gender dysphoria found that during treatment with gender affirming hormones (duration of treatment not reported):

- No participants permanently discontinued treatment.
- No transfemales temporarily discontinued treatment, but 3 transmales temporarily discontinued treatment due to mental health comorbidities (n=2) and androgenic alopecia (n=1). All 3 participants eventually resumed treatment, although timescales were not reported
- No severe complications were reported.
- No transfemales reported minor complications, but 12 transmales developed minor complications which were: severe acne (n=7), androgenic alopecia (n=1), mild dyslipidaemia (n=3) and significant mood swings (n=1).

In children and adolescents with gender dysphoria, what is the cost-effectiveness of gender-affirming hormones compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?

No cost-effectiveness evidence was found for gender-affirming hormones for children and adolescents with gender dysphoria.

From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria?

Some studies reported data separately for the following subgroups of children and adolescents with gender dysphoria:

- Sex assigned at birth males (transfemales).
- Sex assigned at birth females (transmales).
- Tanner stage at which GnRH analogue or gender-affirming hormones started.
- Diagnosis of a mental health condition.

Some direct comparisons of transfemales and transmales were included. No evidence was found for other specified subgroups.

Sex assigned at birth males (transfemales)
Impact on mental health
In the study by Kuper et al. 2020 in 33 to 45 (number varies by outcome) sex assigned at birth males (transfemales) with gender dysphoria found that during treatment with gender-affirming hormones changes were seen in depression, anxiety and anxiety-related symptoms from baseline to follow-up (mean duration of treatment 10.9 months). The authors did not report any statistical analyses, so it is unclear if any changes were statistically significant.

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, suicide risk (measured using the ASQ) is not statistically significant different in transfemales compared with transmales, between baseline and the final assessment at about 12 months (p=0.79).

The study by Achille et al. 2020 in 17 transfemales with gender dysphoria found that during treatment with gender-affirming hormones, suicidal ideation (measured using the PHQ 9_Modified for Teens with additional questions for suicidal ideation) was reported in 11.8% (2/17) of transfemales at baseline compared with 5.9% (1/17) at about 12-months follow-up (no statistical analysis was reported).

Impact on quality of life

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, quality of life (measured using the GWBS of the Paediatric Quality of Life Inventory) was not statistically significant different in transfemales compared with transmales, between baseline and the final assessment at about 12 months (p=0.32).

Bone density

The studies by Klink et al. 2015 and Vlot et al. 2017 provided evidence on bone density in transfemales; see above for details.

Change in clinical parameters

The study by Klaver et al. 2020 provided evidence on the following clinical parameters in transfemales:

- Glucose levels, insulin levels and insulin resistance.
- Total cholesterol, HDL cholesterol and LDL cholesterol and triglycerides.
- Blood pressure.
- Body mass index.

See above for details.

Treatment discontinuation and adverse effects

The study by Khatchadourian et al. 2014 provided evidence on treatment discontinuation and adverse effects in transfemales; see above for details.

Sex assigned at birth females (transmales)

Impact on mental health

In the study by Kuper et al. 2020 in 65 to 78 (number varies by outcome) sex assigned at birth females (transmales) with gender dysphoria found that during treatment with gender-affirming hormones, changes were seen in depression, anxiety and anxiety-related symptoms from baseline to 10.9 month follow-up. The authors did not report any statistical analyses, so it is unclear if any changes were statistically significant.

The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, suicide risk (measured using the ASQ) is not statistically significantly different in transmales compared with transfemales, between baseline and the final assessment (p=0.79).

The study by Achille et al. 2020 in 33 transmales with gender dysphoria found that during treatment with gender-affirming hormones, suicidal ideation (measured using the PHQ 9_Modified for Teens with additional questions for suicidal ideation) was reported in 9.1% (3/33) of transmales at baseline compared with 6.1% (2/33) at about 12-months follow-up (no statistical analysis reported).

Impact on quality of life
The study by Allen et al. 2019 in 47 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, quality of life (measured using the GWBS of the Paediatric Quality of Life Inventory) was not statistically significantly different in transmales compared with transfemales, between baseline and the final assessment at about 12 months (p=0.32).

Bone density
The studies by Klink et al. 2015, Stoffers et al. 2019 and Vlot et al. 2017 provided evidence on bone density in transmales; see above for details.

Change in clinical parameters
The study by Klaver et al. 2020 provided evidence on the following clinical parameters in transmales:

- Glucose levels, insulin levels and insulin resistance.
- Total cholesterol, HDL cholesterol and LDL cholesterol and triglycerides.
- Blood pressure.
- Body mass index.

See above for details.

The study by Stoffers et al. 2019 provided evidence on HbA1c, liver enzymes and renal function in transmales; see above for details.

Treatment discontinuation and adverse effects
The study by Khatchadourian et al. 2014 provided evidence on treatment discontinuation and adverse effects in transmales; see above for details.

Tanner stage at which GnRH analogues or gender-affirming hormones started
The study by Kuper et al. 2020 stated that the impact of Tanner stage on outcomes was considered, but it is unclear if this refers to Tanner stage at the initial assessment, at the start of GnRH analogue treatment or another timepoint. No results were reported.

Diagnosis of a mental health condition
Impact on mental health
The study by Achille et al. 2020 in 50 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, there was no statistically significant difference in depression (measured using the CESD-R and PHQ 9_Modified for Teens) when the results were adjusted for engagement in counselling and medicines for mental health problems, from baseline to about 12-months follow-up.

Impact on quality of life
The study by Achille et al. 2020 in 50 adolescents with gender dysphoria found that during treatment with gender-affirming hormones, there was no statistically significant difference in quality of life (measured using the QLES-Q-SF) when the results were adjusted for engagement in counselling and medicines for mental health problems, from baseline to about 12-months follow-up.

From the evidence selected,

(a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?
(b) what were the ages at which participants commenced treatment with gender-affirming hormones?
(c) what was the duration of treatment with GnRH analogues?

The most commonly reported diagnostic criteria for gender dysphoria was the DSM criteria in use at the time (5/10 studies). In 3 studies (Klaver et al. 2020, Klink et al. 2015 and Vlot et al. 2017) DSM-IV-TR criteria was used. In 2 studies (Kuper et al. 2020 and Stoffers et al. 2019) DSM-V criteria was used. One study from Finland (Kaltiala et al. 2020) used the ICD-10 diagnosis of 'transexualism'. It was not reported how gender dysphoria was defined in the remaining 4 studies.

In the studies, treatment with gender-affirming hormones started at about 16 to 17 years, with a range of about 14 to 19 years. Most studies did not report the duration of treatment with GnRH analogues, but where this was reported there was a wide variation ranging from a few months up to about 5 years (Klaver et al. 2020, Klink et al. 2015 and Stoffers et al. 2019).

Discussion

The key limitation to identifying the effectiveness and safety of gender-affirming hormones for children and adolescents with gender dysphoria is the lack of reliable comparative studies.

All the studies included in the evidence review are uncontrolled observational studies, which are subject to bias and confounding and were of very low certainty using modified GRADE. A fundamental limitation of all the uncontrolled studies included in this review is that any changes in scores from baseline to follow-up could be attributed to a regression-to-the-mean.

The included studies have relatively short follow-up, with an average duration of treatment with gender-affirming hormones between around 1 year and 5.8 years. Further studies with a longer follow-up are needed to determine the long-term effect of gender-affirming hormones for children and adolescents with gender dysphoria.

Most studies included in this review did not report comorbidities (physical or mental health) and no study reported concomitant treatments in detail. Because of this it is not clear whether any changes seen were due to gender-affirming hormones or other treatments the participants may have received.

There is a degree of indirectness in some studies, with some participants included that fall outside of the population of this evidence review. Furthermore, participant numbers are poorly reported in some studies, with high numbers lost to follow-up or outcomes not reported for some participants. The authors provide no explanation for this incomplete reporting.

Details of the gender-affirming hormone treatment regimen are poorly reported in most of the included studies, with limited information provided about the medicines, doses and routes of administration used. It is not clear whether the interventions used in the studies are reflective of current UK practice for children and adolescents with gender dysphoria.

It is difficult to draw firm conclusions for many of the effectiveness and safety outcomes reported in the included studies because many different scoring tools and methods were used to assess the same outcome, often with conflicting results. In addition to this, most outcomes reported across the included studies do not have an accepted minimal clinically important difference (MCID), making it difficult the determine whether any statistically significant changes seen are clinically meaningful. However, the authors of some studies report thresholds to interpret the results of the scoring tools (for example, by linking scores to symptom severity), so some conclusions can be made.

Conclusion

Any potential benefits of gender-affirming hormones must be weighed against the largely unknown long-term safety profile of these treatments in children and adolescents with gender dysphoria.

Results from 5 uncontrolled, observational studies suggest that, in children and adolescents with gender dysphoria, gender-affirming hormones are likely to improve symptoms of gender dysphoria, and may also improve depression, anxiety, quality of life, suicidality, and psychosocial functioning. The impact of treatment on body image is unclear. All results were of very low certainty using modified GRADE.

Safety outcomes were reported in 5 observational studies. Statistically significant increases in some measures of bone density were seen following treatment with gender-affirming hormones, although results varied by bone region (lumber spine versus femoral neck) and by population (transfemales versus transmales). However, z-scores suggest that bone density remained lower in transfemales and transmales compared with an equivalent cisgender population. Results from 1 study of gender-affirming hormones started during adolescence reported statistically significant increases in blood pressure and body mass index, and worsening of the lipid profile (in transmales) at age 22 years, although longer term studies that report on cardiovascular event rates are required. Adverse events and discontinuation rates associated with gender-affirming hormones were only reported in 1 study, and no conclusions can be made on these outcomes.

This review did not identify sub-groups of patients who may benefit more from gender-affirming hormones.

No cost-effectiveness evidence was found to determine whether gender-affirming hormones are a cost-effective treatment for children and adolescents with gender dysphoria.

3. Methodology

Review questions

The review question(s) for this evidence review are:

1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

2. For children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
3. For children and adolescents with gender dysphoria, what is the cost-effectiveness of gender-affirming hormones compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?
4. From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria?
5. From the evidence selected,
   (a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?
   (b) what were the ages at which participants commenced treatment with gender-affirming hormones?
   (c) what was the duration of GnRH analogues treatment?

See appendix A for the full review protocol.

Review process

The methodology to undertake this review is specified by NHS England in their 'Guidance on conducting evidence reviews for Specialised Services Commissioning Products' (2020).

The searches for evidence were informed by the PICO and were conducted on 21 July 2020.

See appendix B for details of the search strategy.

Results from the literature searches were screened using their titles and abstracts for relevance against the criteria in the PICO framework. Full text references of potentially relevant evidence were obtained and reviewed to determine whether they met the inclusion criteria for this evidence review.

See appendix C for evidence selection details and appendix D for the list of studies excluded from the review and the reasons for their exclusion.

Relevant details and outcomes were extracted from the included studies and were critically appraised using a checklist appropriate to the study design. See appendix E and appendix F for individual study and checklist details.

The available evidence was assessed by outcome for certainty using modified GRADE. See appendix G for GRADE Profiles.

4. Summary of included studies

Ten observational studies were included in the evidence review. Seven studies were retrospective observational studies (Allen et al. 2019, Kaltiala et al. 2020, Khatchadourian et al. 2014, Klaver et Al. 2020, Klink et al. 2015, Stoffers et al. 2019, Vlot et al. 2017) and three studies were prospective longitudinal observational studies (Achille et al. 2020, Kuper et al. 2020, Lopez de Lara et al. 2020).

The terminology used in this topic area is continually evolving and is different depending on stakeholder perspectives. In this evidence review we have used the phrase 'people's assigned sex at birth' rather than saying natal or biological sex and 'cross sex hormones' are now referred to as 'gender-affirming hormones'. The research studies may use historical terms which are no longer considered appropriate.

Table 1 provides a summary of these included studies and full details are given in appendix E.

Table 1 Summary of included studies

Study	Population	Intervention and comparison	Outcomes reported
Achille et al. 2020 Prospective longitudinal study Single centre, New York, United States	50 children, adolescents and young adults with gender dysphoria; 17 transfemales and 33 transmales Mean age at baseline was 16.2 years (SD 2.2)	**Intervention** Endocrine interventions (the collective term used for puberty suppression and gender-affirming hormones) were introduced as per Endocrine Society and the World Professional Association for Transgender Health (WPATH) guidelines Puberty suppression was: • GnRH analogue and/or anti-androgens (transfemales) • GnRH analogue or medroxyprogesterone (transmales) Once eligible, gender-affirming hormones were offered, these were: • Oestradiol (transfemales)	**Critical Outcomes** *Impact on mental health* • Depression- The Center for Epidemiologic Studies Depression Scale (CESD-R) • Depression- The Patient Health Questionnaire Modified for Teens (PHQ 9_Modified for Teens) *Impact on quality of life* • Quality of Life Enjoyment and Satisfaction Questionnaire (QLES-Q-SF) **Important Outcomes** *None reported*

Study	Population	Intervention and comparison	Outcomes reported
		• Testosterone (transmales) Doses and formulations not reported After about 12-months treatment ('wave 3'): • 24 people (48%) were on gender-affirming hormones alone • 12 people (24%) were on puberty suppression alone • 11 people (22%) were on both gender-affirming hormones and puberty suppression • 3 people (6%) were on no endocrine intervention **Comparison** No comparison group. Change over time reported	
Allen et al. 2019 Retrospective longitudinal study Single centre, Kansas City, USA	47 adolescents and young adults with gender dysphoria: 14 transfemales and 33 transmales Mean age at administration (start of treatment) 16.5 years	**Intervention** 39 participants received gender-affirming hormones only 8 participants received hormones and a GnRH analogue Mean duration of treatment with gender-affirming hormones was 349 days (range 113 to 1,016) **Comparison** No comparison group. Comparison over time reported	**Critical Outcomes** *Impact on mental health* • Suicidality- Ask Suicide-Screening Questions (ASQ) instrument *Impact on quality of life* • General Well-Being Scale (GWBS) of the Pediatric Quality of Life Inventory **Important Outcomes** *None reported*
Kaltiala et al. 2020	52 adolescents with gender dysphoria: 11 transfemales and 41 transmales.	**Intervention** Hormonal sex assignment treatment – details of	**Critical Outcomes** *Impact on mental health*

Study	Population	Intervention and comparison	Outcomes reported
Retrospective chart review Single centre, Tampere, Finland	Mean age at diagnosis 18.1 years (range 15.2 to 19.9)	intervention not reported, although all patients received gender-affirming hormones. **Comparison** No comparison group. Comparison over time reported	• Need for mental health treatment **Important Outcomes** *Psychosocial Impact* Measure of functioning in different domains of adolescent development, which were: • Living with parent(s)/ guardians • Normative peer contacts • Progresses normatively in school/ work • Has been dating or had steady relationships • Is age-appropriately able to deal with matters outside of the home
Khatchadourian et al. 2014 Retrospective chart review Single centre, Vancouver, Canada	84 young people with gender dysphoria, of whom 63 received gender-affirming hormones. Median age at start of gender-affirming hormones was: • 17.3 years (range 13.7-19.8) for testosterone • 17.9 years (range 13.3-22.3) for oestrogen	**Intervention** Transfemales: Oestrogen (oral micronized 17β-oestradiol) Transmales: Testosterone (injectable testosterone enanthate and/or cypionate) 19 participants (30%) had previously received a GnRH analogue **Comparison** No comparison group. Comparison over time reported.	**Critical Outcomes** *None reported* **Important Outcomes** *Safety:* • Adverse events • Discontinuation rates
Klaver et al. 2020 Retrospective chart review Single centre, Amsterdam, Netherlands	192 people with gender dysphoria who started GnRH analogues before the age of 18 years, and started gender-affirming hormones within 1.5 years of their 22nd birthday.	**Intervention** Oral oestrogen or intramuscular (IM) testosterone **Comparison**	**Critical Outcomes** *None reported* **Important Outcomes** *Safety* • Body mass index (BMI)

Study	Population	Intervention and comparison	Outcomes reported
	Mean age at start of gender-affirming hormones: • Transfemale – 16.4 years (SD 1.1) • Transmale – 16.9 years (SD 1.9)	No comparison group. Comparison over time reported	• Systolic blood pressure • Diastolic blood pressure • Glucose • Insulin • HOMA-IR • Total cholesterol • HDL cholesterol • LDL cholesterol • Triglycerides
Klink et al. 2015 Retrospective longitudinal study Single centre, Amsterdam, Netherlands	34 young people with gender dysphoria who had received GnRH analogues, gender-affirming hormones and gonadectomy. The study included 15 transfemales and 19 transmales; mean age at start of gender-affirming hormones was 16.6 years (SD 1.4) and 16.4 years (SD 2.3) respectively. At the start of gender-affirming hormone treatment, in the transfemale subgroup the median Tanner P was 4 (IQR 2) and the median Tanner G was 12 (IQR 11) In the transmale subgroup the median Tanner B was 5 (IQR 2) and the median Tanner P was 5 (IQR 0)	**Intervention** Transfemales – oral 17-β oestradiol (incremental dosing) Transmales – IM testosterone (Sustanon 250 mg/ml; incremental dosing) Median duration of treatment with gender-affirming hormones for transfemales was 5.8 years (range 3.0 to 8.0) and for transmales was 5.4 years (range 2.8 to 7.8) The GnRH analogue was subcutaneous (SC) triptorelin 3.75 mg every 4 weeks No details of gonadectomy reported **Comparison** No comparison group. Comparison over time reported.	**Critical Outcomes** None **Important Outcomes** *Safety* • Bone mineral apparent density (BMAD) • Bone mineral density (BMD) Measures reported at 3 timepoints: start of GnRH analogue treatment, start of gender-affirming hormone treatment and age 22 years.
Kuper et al. 2020 Prospective longitudinal study	Children and adolescents with gender dysphoria (9 to18 years), n=148, of whom: • 25 received puberty suppression only	**Intervention** Gender-affirming hormones, guided by Endocrine Society Clinical Practice Guidelines	**Critical Outcomes** *Impact on mental health* • Depression- Quick Inventory of Depressive

Study	Population	Intervention and comparison	Outcomes reported
Single centre, Texas, USA	• 93 received gender-affirming hormone therapy only • 30 received both Mean age 14.9 years	**Comparison** No comparison group. Comparison over time reported.	Symptoms (QIDS), self-reported • Depression- QIDS, clinician-reported • Anxiety- Screen for Child Anxiety Related Emotional Disorders (SCARED) • Panic- specific questions from SCARED • Generalised anxiety-specific questions from SCARED • Social anxiety - specific questions from SCARED • Separation anxiety-specific questions from SCARED • School avoidance-specific questions from SCARED **Important Outcomes** *Impact on body image* • Body Image Scale (BIS)
Lopez de Lara et al. 2020 Prospective analytical study Single centre, Madrid, Spain	23 adolescents with gender dysphoria: 7 transfemales and 16 transmales. Mean age at baseline was 16 years (range 14 to 18)	**Intervention** Gender-affirming hormones: • Oral oestradiol • Intramuscular testosterone Participants had previously received GnRH analogues in the intermediate pubertal stages (Tanner 2 to 3). Participants were assessed twice: • pre-treatment (T0), • after 12 months treatment with gender-affirming hormones (T1)	**Critical Outcomes** *Impact on gender dysphoria* • Utrecht Gender Dysphoria Scale (UGDS) *Impact on mental health* • Depression- Beck Depression Inventory II (BDI-II) • Anxiety- State-Trait Anxiety Inventory **Important Outcomes** *Psychosocial Impact* • Family functioning-Family APGAR test • Patient strengths and difficulties-Strengths and Difficulties Questionnaire,

Study	Population	Intervention and comparison	Outcomes reported
		Comparison No comparison group. Comparison over time reported.	Spanish Version (SDQ-Cas).
Stoffers et al. 2019 Retrospective chart review Single centre, Leiden, Netherlands	62 transmales with gender dysphoria. Patients had received a GnRH analogue and more than 6 months of testosterone treatment. Median age at start of testosterone was 17.23 years (range 14.9 to 18.4) Median treatment duration was 12 months (range 5 to 33) Change over time	**Intervention** Testosterone intramuscular injections (Sustanon 250 mg). Dose was titrated to a maintenance dose of 125 mg every 2 weeks. Participants who started GnRH analogues at 16 years or older had their dose increased more rapidly. Some participants chose to receive testosterone every 3-4 weeks, and participants could switch to transdermal preparations if needed. **Comparison** No comparison group. Comparison over time reported.	**Critical Outcomes** None **Important Outcomes** *Safety* • Body mass index (BMI) • Blood pressure • BMD • Acne • Liver enzymes • Creatinine • Urea • HbA1c
Vlot et al. 2017 Retrospective chart review Single centre, Amsterdam, Netherlands	70 children and adolescents with gender dysphoria Median age at baseline – • 13.5 years (11.5-18.3) for transfemales • 15.1 years (range 11.7-18.6) for transmales Comparison is change over time. 24 month follow-up.	**Intervention** Oestrogen or testosterone (had previously received triptorelin for puberty suppression) **Comparison** No comparison group. Comparison over time reported.	**Critical Outcomes** None **Important Outcomes** *Safety* • Bone mineral apparent density (BMAD)

5. Results

In children and adolescents with gender dysphoria, what is the clinical effectiveness of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement
Clinical Effectiveness	

Critical outcomes	
Impact on gender dysphoria **Certainty of evidence: very low**	This is a critical outcome because gender dysphoria in children and adolescents is associated with significant distress and problems with functioning. One uncontrolled, prospective, observational study (Lopez de Lara et al. 2020) provided evidence relating to the impact on gender dysphoria, measured using the Utrecht Gender Dysphoria Scale (UGDS) score during the first year of treatment with gender-affirming hormones. The UGDS is a validated, screening tool for both adolescents and adults, used to assess gender dysphoria. It consists of 12 items, to be answered on a 1- to 5-point scale, resulting in a sum score between 12 and 60. The authors state that the cut-off point to identify gender dysphoria is 40 points. The higher the UGDS score the greater the gender dysphoria. In this study (n=23), the mean (±SD) UGDS score was statistically significantly reduced (improved) from 57.1 (±4.1) points at baseline to 14.7 points (±3.2) at 12 months (p<0.001). A UGDS score below 40 suggests an absence of gender dysphoria **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly improve gender dysphoria from baseline to 12 months follow-up. The mean UGDS score was below the threshold for gender dysphoria at follow-up.**
Impact on mental health: depression **Certainty of evidence: very low**	This is a critical outcome because depression may impact on social, occupational, or other areas of functioning in children and adolescents. Four observational studies (Achille et al. 2020; Kaltiala et al. 2020; Kuper et al. 2020; Lopez de Lara et al. 2020) provided evidence relating to the impact on depression in children and adolescents with gender dysphoria, with follow-up of around 12 months. Five different outcome measures for depression were reported. **Beck Depression Inventory (BDI-II)** One uncontrolled, prospective, analytical study (Lopez de Lara et al. 2020) reported the change in BDI-II. The BDI-II is a valid, reliable, and widely used tool for assessing depressive symptoms. There are no specific scores to categorise depression severity, but it is suggested that 0 to 13 is minimal symptoms, 14 to 19 is mild depression, 20 to 28 is moderate depression, and severe depression is 29 to 63. In Lopez de Lara et al. 2020 (n=23) the mean (±SD) BDI-II score was statistically significantly reduced (improved) from 19.3 (±5.5) points at baseline to 9.7 (±3.9) points at 12 months (p<0.001) **(VERY LOW)**. **Center for Epidemiologic Studies Depression (CESD-R)** One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in CESD-R scale. The CESD-R is a valid, widely used tool to assess depressive symptoms. Total score ranges from 0 to 60, with higher scores indicating more depressive symptoms. There are no specific scores to categorise depression severity, although the authors of the study suggest that a total CESD-R score less than 16 suggests no clinical depression.

	In Achille et al. 2020 (n=50), the mean CESD-R score statistically significantly reduced (improved) from 21.4 points at baseline to 13.9 points at about 12 months follow-up (p<0.001; standard deviation not reported) **(VERY LOW)**. **Patient Health Questionnaire (PHQ 9) Modified for Teens** One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in PHQ 9_Modified for Teens score. The PHQ 9_Modified for Teens is a validated tool to assess depression, dysthymia and suicide risk. The tool consists of 9 questions scored from 0 to 3 (total score 0 to 27), plus an additional 4 questions that are not scored. A score of 0 to 4 suggests no or minimal depressive symptoms, 5 to 9 mild, 10 to 14 moderate, 15 to 19 moderately severe, and 20-27 severe symptoms. In Achille et al. 2020 (n=50), the mean PHQ 9_Modified for Teens score statistically significantly reduced (improved) from baseline to around 12 months follow-up, although absolute scores were not reported numerically (p<0.001). From the visual representation of results, the PHQ-9_Modified for Teens score is about 9 at baseline and about 5 at final follow-up **(VERY LOW)**. **Quick Inventory of Depressive Symptoms (QIDS)** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported the change in QIDS, clinician-reported and self-reported. Both the clinician-reported and self-reported QIDS are validated tools to assess depressive symptoms. The tool consists of 16 items, with the highest score for 9 domains (sleep, weight, psychomotor changes, depressed mood, decreased interest, fatigue, guilt, concentration, and suicidal ideation) added to give a total score ranging from 0 to 27. A score of 0 to 5 suggests no depression, 6 to 10 mild symptoms, 11 to 15 moderate symptoms, 16 to 20 severe symptoms, and 21 to 27 very severe symptoms. In Kuper et al. 2020 (n=105), the mean (±SD) QIDS self-reported score was 9.6 points (±5.0) at baseline and 7.4 (±4.5) after 10.9 months of treatment with gender-affirming hormones (no statistical analysis reported). The mean (±SD) QIDS clinician-reported score was 5.9 points (±4.1) at baseline and 6.0 (±3.8) after 10.9 months of treatment with gender-affirming hormones (no statistical analysis was reported) **(VERY LOW)**. **Participants needing treatment for depression** One observational study (Kaltiala et al. 2020) reported the proportion of participants needing treatment for depression before or during the initial assessment and during the 12-month follow-up period after starting gender-affirming hormones. In Kaltiala et al. 2020 (n=52), statistically significantly fewer participants needed treatment for depression during the 12-month 'real life' phase (15%, 8/52) compared with before or during the assessment (54%, 28/52; p<0.001). No details of what treatments for depression the participants received are reported **(VERY LOW)**.

	These studies provide very low certainty evidence that during treatment with gender-affirming hormones depression is reduced from baseline to about 12 months follow-up. However, most participants had mild symptoms at the start of treatment.
Impact on mental health: anxiety **Certainty of evidence: very low**	This is a critical outcome because anxiety may impact on social, occupational, or other areas of functioning in children and adolescents. Three observational studies (Kaltiala et al. 2020; Kuper et al. 2020; Lopez de Lara et al. 2020) provided evidence relating to the impact on anxiety in children and adolescents with gender dysphoria. **State-Trait Anxiety Inventory (STAI)** One uncontrolled, prospective, analytical study (Lopez de Lara et al. 2020) reported the change in STAI scores. STAI is a validated and commonly used measure of trait and state anxiety. It has 20 items and can be used in clinical settings to diagnose anxiety and to distinguish it from depressive illness. Higher scores indicate greater anxiety. In Lopez de Lara et al. 2020 (n=23), the mean (±SD) STAI-State subscale was statistically significantly reduced (improved) with gender-affirming hormones from 33.3 points (±9.1) at baseline to 16.8 points (±8.1) at 12 months (p<0.001). The mean STAI-Trait subscale scores also statistically significantly reduced (improved) from 33.0 points (±7.2) at baseline to 18.5 points (±8.4) at 12 months (p<0.001) **(VERY LOW)**. **Screen for Child Anxiety Related Emotional Disorders (SCARED)** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported anxiety symptoms using the SCARED questionnaire. Other anxiety-related symptoms using specific questions from the SCARED questionnaire were also reported: panic, generalised anxiety, social anxiety, separation anxiety and school avoidance. SCARED is a validated, 41-point questionnaire, with each item scored 0 to 2. A total score of 25 or more is suggestive of anxiety disorder, with scores above 30 being more specific. Certain scores for specific questions may indicate the presence of other anxiety-related disorders: • A score of 7 or more in questions related to panic disorder or significant somatic symptoms may indicate the presence of these. • A score of 9 or more in questions related to generalised anxiety disorder may indicate the presence of this. • A score of 5 or more in questions related to separation anxiety may indicate the presence of this. • A score of 8 or more in questions related to social anxiety disorder may indicate the presence of this. • A score of 3 or more in questions related to significant school avoidance may indicate the presence of this. In Kuper et al. 2020 (n=80 to 82, varies by outcome), small reductions were seen in anxiety, panic, generalised anxiety, social anxiety and separation anxiety and school avoidance symptoms (measured using the SCARED questionnaire) from baseline to follow-up (mean duration of treatment 10.9 months). The statistical significance of these findings are unknown as no statistical analyses were reported **(VERY LOW)**.

	Participants needing treatment for anxiety One observational study (Kaltiala et al. 2020) reported the proportion of participants needing treatment for anxiety before or during initial assessment and during the 12-month follow-up period after starting gender-affirming hormones. In Kaltiala et al. 2020 (n=52), statistically significantly fewer participants needed treatment for anxiety during the 12-month 'real life' phase (15%, 8/52) compared with before or during the assessment (48%, 25/52; p<0.001). No details of what treatments for anxiety the participants received are reported **(VERY LOW)**. **These studies provide very low certainty evidence that during treatment with gender-affirming hormones anxiety symptoms may be reduced from baseline to around 12 months follow-up.**
Impact on mental health: suicidality and self-injury **Certainty of evidence: very low**	These are critical outcomes because self-harm and thoughts of suicide have the potential to result in significant physical harm and, for completed suicides, the death of the young person. Four observational studies (Achille et al. 2020; Allen et al. 2019; Kaltiala et al. 2020; Kuper et al. 2020) provided evidence relating to suicidal ideation in children and adolescents with gender dysphoria, with an average follow-up of around 12 months. **Ask Suicide-Screening Questions (ASQ)** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in ASQ. This is a 4-item dichotomous (yes/no) response measure designed to identify risk of suicide. The authors of Allen et al. 2019 amended 1 question in the ASQ ("*Have you ever tried to kill yourself?*") by prefacing it with "*In the past few weeks . . .*" as they were not investigating lifetime incidence. A response of 'no' is scored as 0 and a response of 'yes' is scored as 1; each item is summed to give an overall score for suicidal ideation ranging from 0 to 4. A person is considered to have screened positive if they answer 'yes' to any item with higher scores indicating higher levels of suicidal ideation. In Allen et al. 2019 (n=39), the adjusted mean (±SE) ASQ score statistically significantly reduced from 1.11 points (±0.22) at baseline to 0.27 points (±0.12) after a mean duration of treatment of about 12 months (p<0.001) **(VERY LOW)**. **PHQ 9_Modified for Teens (additional questions for suicidal ideation)** One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in suicidal ideation measured using additional questions from the PHQ 9_Modified for Teens. This is a validated tool to assess depression, dysthymia and suicide risk (see above for detailed description). In addition to the 9 scored questions, the PHQ 9_Modified Teens asked 4 additional questions relating to suicidal ideation and difficulty dealing with problems of life. Responses to the PHQ 9_Modified for Teens were used to determine if the participant had suicidal ideation or not, but specific details of how this was determined are not reported.

	In Achille et al. 2020 (n=50), 10% (5/50) of participants had suicidal ideation at baseline and 6% (3/50) had suicidal ideation after about 12 months treatment with gender-affirming hormones (no statistical analysis reported) **(VERY LOW)**. **Suicidality and non-suicidal self-injury** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported on suicidal ideation, suicide attempts and non-suicidal self-injury, although it was unclear how and when this outcome was measured. In Kuper et al. 2020 (n=130), 25% of participants reported suicidal ideation 1 month before the initial assessment and 38% reported this during the follow-up period (no statistical analysis reported). Suicide attempts were reported in 2% of participants at 3 months before the initial assessment and 5% during follow-up. Self-injury was reported in 10% of participants at 3 months before the initial assessment and 17% during follow-up. No statistical analysis was reported for any outcomes. Mean duration of gender-affirming hormone treatment was 10.9 months **(VERY LOW)**. **Participants needing treatment for suicidality or self-harm** One observational study (Kaltiala et al. 2020) reported the proportion of participants requiring treatment for suicidality or self-harm before or during initial assessment and during the 12-month follow-up period after starting gender-affirming hormones. In Kaltiala et al. 2020 (n=52) statistically significantly fewer participants needed treatment for suicidality or self-harm during the 12-month 'real life' phase (4%, 2/52) compared with before or during the assessment (35%, 18/52; p<0.001). No details of what treatments for suicidal ideation or self-harm the participants received are reported **(VERY LOW)**. **These studies provide very low certainty evidence that gender-affirming hormones may reduce suicidality from baseline to about 12 months follow-up. However, results are inconsistent and it is difficult to draw conclusions.**
Impact on mental health: other **Certainty of evidence: very low**	This is a critical outcome because mental health problems may impact on social, occupational, or other areas of functioning in children and adolescents. One observational study (Kaltiala et al. 2020) reported the proportion of participants needing treatment for either psychotic symptoms or psychosis, substance abuse, autism, attention deficit hyperactivity disorder (ADHD) or eating disorders before or during initial assessment and during the 12-month follow-up period after starting gender-affirming hormones. In Kaltiala et al. 2020 (n=52) there was no statistically significant difference in the number of people needing treatment for either psychotic symptoms / psychosis, substance abuse, autism, attention deficit hyperactivity disorder (ADHD) or eating disorders during the 12-month 'real life' phase compared with before or during the assessment.

	No details of which specific treatments the participants received are reported **(VERY LOW)**. **This study provides very low certainty evidence on the need for treatment for either psychotic symptoms or psychosis, conduct problems or antisocial behaviour, substance abuse, autism, attention deficit hyperactivity disorder (ADHD) or eating disorders during treatment with gender-affirming hormones. No conclusions could be drawn.**
Impact on quality of life score **Certainty of evidence: very low**	This is a critical outcome because gender dysphoria in children and adolescents may be associated with a significant reduction in health-related quality of life. Two uncontrolled longitudinal studies Achille et al. 2020; Allen et al. 2019) provided evidence relating to quality of life in children and adolescents with gender dysphoria. **Quality of Life Enjoyment and Satisfaction Questionnaire (QLES-Q-SF)** One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in QLES-Q-SF scores from baseline to about 12 months of treatment with gender-affirming hormones. QLES-Q-SF is a validated questionnaire, consisting of 15 questions that rate quality of life on a scale of 1 (poor) to 5 (very good). In Achille et al. 2020 (n=50), the mean QLES-Q-SF score was statistically significantly reduced from baseline to about 12 months ($p<0.001$). However, absolute scores are not reported numerically **(VERY LOW)**. **General Well-Being Scale (GWBS) of the Paediatric Quality of Life Inventory** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in adjusted mean GWBS of the Paediatric Quality of Life Inventory score from baseline to about 12 months of treatment with gender-affirming hormones. The GWBS of the Paediatric Quality of Life Inventory contains 7 items that measure two dimensions: general wellbeing (6 items) and general health (1 item). Each item is scored from 0 to 4, and the total score is linearly transformed to a 0 to 100 scale. Higher scores reflect fewer perceived problems and greater well-being. In Allen et al. 2019 (n=47), the adjusted mean (±SE) GWBS of the Paediatric Quality of Life Inventory score was statistically significantly increased (improved) from 61.70 (±2.43) points at baseline to 70.23 (±2.15) points at about 12 months ($p<0.002$) **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly improve quality of life and well-being from baseline to 12 months follow-up.**
Important outcomes	
Impact on body image	This is an important outcome because some children and adolescents with gender dysphoria may want to take steps to suppress features of

Certainty of evidence: very low	their physical appearance associated with their sex assigned at birth or accentuate physical features of their desired gender. One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) provided evidence relating to the impact on body image in children and adolescents with gender dysphoria who started treatment with gender-affirming hormones (median duration 10.9 months; range 1 to 18), measured by the change in Body Image Scale (BIS) score. BIS is a validated 30-item scale covering 3 aspects: primary, secondary and neutral body characteristics. Higher scores represent a higher degree of body dissatisfaction. In Kuper et al. 2020 (n=86), the mean (±SD) BIS score was 70.7 points (±15.2) at baseline and 51.4 points (±18.3) at follow-up (no statistical analysis reported) **(VERY LOW)**. **This study provides very low certainty evidence on the effects of gender-affirming hormones on body image during treatment with gender-affirming hormones (mean duration of treatment 10.9 months). No conclusions could be drawn.**
Psychosocial impact **Certainty of evidence: very low**	This is an important outcome because gender dysphoria in children and adolescents is associated with internalising and externalising behaviours, and emotional and behavioural problems which may impact on social and occupational functioning. Two uncontrolled, observational studies (Kaltiala et al. 2020; Lopez de Lara et al. 2020) provided evidence related to psychosocial impact in children and adolescents with gender dysphoria. **Family APGAR (Adaptability, Partnership, Growth, Affection and Resolve) test** One uncontrolled, prospective, analytical study (Lopez de Lara et al. 2020) reported the Family APGAR test. The Family APGAR test is a 5-item questionnaire, with higher scores indicating better family functioning. The authors reported the following interpretation of the test: functional, 17 to 20 points; mildly dysfunctional, 16 to 13 points; moderately dysfunctional, 12 to 10 points; severely dysfunctional, <9 points. In Lopez de Lara et al. 2020 (n=23), the mean Family APGAR test score was unchanged from baseline (17.9 points) to 12-month follow-up (18.0 points; no statistical analysis or standard deviations reported) **(VERY LOW)**. **Strengths and Difficulties Questionnaire (SDQ)** One uncontrolled, prospective, analytical study (Lopez de Lara et al. 2020) reported on behaviour using the Strengths and Difficulties Questionnaire (SDQ, Spanish version). The SDQ includes 25-items covering emotional symptoms, conduct problems, hyperactivity/ inattention, peer relationship problems and prosocial behaviour. The authors state that a score of more than 20 suggests having a behavioural disorder (normal 0 to 15, borderline 16 to 19, abnormal 20 to 40).

	In Lopez de Lara et al. 2020 (n=23), the mean (±SD) SDQ score was statistically significantly reduced (improved) from 14.7 points (±3.3) at baseline to 10.3 points (±2.9) at 12-month follow-up (p<0.001) **(VERY LOW)**. **Psychosocial functioning** One uncontrolled, retrospective chart review (Kaltiala et al. 2020) reported various markers of functioning in adolescent development, covering living arrangements, peer contacts, school or work progress, relationships, and ability to cope with matters outside the home. These measures were reported during the gender identity assessment and at about 12 months after starting gender-affirming hormones (referred to as the 'real-life phase'). In Kaltiala et al. 2020 (n=52), from the gender identity assessment to the 12-month follow-up period: • statistically significantly fewer participants were living with parents or guardians (73% versus 40%, p=0.001) • statistically significantly fewer participants had normal peer contacts (89% versus 81%, p<0.001) • there was no statistically significant difference in progress in school or work (64% versus 60%, p=0.69) • there was no statistically significant difference in the number of participants who had been dating or in steady relationships (62% versus 58%, p=0.51) • there was no statistically significant difference in the participant's ability to cope with matters outside of the home (81% versus 81%, p=1.00) **(VERY LOW)**. **These studies provide very low certainty evidence that gender-affirming hormones statistically significantly improve behavioural problems (measured by SDQ score). However, the SDQ score was in the 'normal' range at baseline and at 12-month follow up. There was no significant impact on other measures of psychosocial functioning.**
Engagement with health care services	This is an important outcome because patient engagement with health care services will impact on their clinical outcomes. No evidence was identified.
Impact on extent of and satisfaction with surgery	This is an important outcome because some children and adolescents with gender dysphoria may proceed to transitioning surgery. No evidence was identified.
De-transition	This is an important outcome because there is uncertainty about the short- and long-term safety and adverse effects of gender-affirming hormones in children and adolescents with gender dysphoria No evidence was identified.

Abbreviations: APGAR: Adaptability, Partnership, Growth, Affection and Resolve; ASQ: Ask Suicide-Screening Questions; BDI-II: Beck Depression Inventory II; BIS: Body Image Scale; CESD-R: Center for Epidemiologic Studies Depression; GWBS: General Well-Being Scale; p: p-value; PHQ 9_Modified for Teens: Patient Health Questionnaire Modified for Teens; QIDS: Quick Inventory of Depressive Symptoms; QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire; SCARED: Screen for Child Anxiety Related Emotional Disorders;

SD: standard deviation; SE: standard error; SDQ: Strengths and Difficulties Questionnaire; STAI: State-Trait Anxiety Inventory; UGDS: Utrecht Gender Dysphoria Scale.

In children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement
Safety	
Change in bone density: lumbar spine **Certainty of evidence: very low**	This is an important outcome because childhood and adolescence is a key time for bone development and gender-affirming hormones may affect bone development, as shown by changes in lumbar spine bone density. Three uncontrolled, observational studies (2 retrospective and 1 prospective) provided evidence related to bone density: lumbar spine in children and adolescents with gender dysphoria. This was reported as either bone mineral density (BMD), bone mineral apparent density (BMAD), or both. One study reported change in bone density from start of treatment with gender-affirming hormones to age 22 years (Klink et al. 2015). Two studies reported change in bone density from start of gender-affirming hormones up to 24-month follow-up (Stoffers et al. 2019 and Vlot et al. 2017). All participants had previously been treated with a GnRH analogue. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. **Bone mineral apparent density (BMAD)** Two uncontrolled, observational studies reported change in lumbar BMAD (Klink et al. 2015; Vlot et al. 2017). BMAD is a size adjusted value of BMD, incorporating bone size measurements using a UK reference population of growing cis-gender adolescents (up to age 17 years). BMAD is used to correct for height and height gain and may provide a more accurate estimate of bone density in growing adolescents. BMAD was reported as g/cm^3 and as z-scores. Z-scores report how many standard deviations from the mean a measurement sits. A z-score of 0 is equal to the mean, a z-score of -1 is equal to 1 standard deviation below the mean, and a z-score of +1 is equal to 1 standard deviation above the mean. A cis-gender population was used to calculate the bone density z-score, meaning transfemales were compared with cis-males and transmales were compared with cis-females. In Klink et al. 2015 (n=34): • There was no statistically significant difference in lumbar spine BMAD z-score from starting gender-affirming hormones to age 22 years in transfemales. • The z-score for lumbar spine BMAD was statistically significantly higher at age 22 years compared with the start of gender-affirming hormones in transmales (z-score [±SD]: start of hormones -0.50 [±0.81], age 22 years -0.033 [±0.95], p=0.002).

- Actual lumbar spine BMAD values in g/cm^3 were statistically significantly higher at age 22 years compared with the start of gender-affirming hormones in transfemales and transmales **(VERY LOW)**.

In Vlot et al. 2017 (n=70):

- The z-score for lumbar spine BMAD in transfemales with a bone age of <15 years was statistically significantly higher at 24-month follow-up compared with start of gender-affirming hormones (z-score [range]: start of hormones -1.52 [-2.36 to 0.42], 24-month follow-up -1.10 [-2.44 to 0.69], p≤ 0.05). Statistically significant improvements in z-score for lumbar spine BMAD in transfemales with a bone age of ≥15 years were also seen (z-score [range]: start of hormones -1.15 [-2.21 to 0.08], 24-month follow-up -0.66 [-1.66 to 0.54], p≤ 0.05).
- The z-score for lumbar spine BMAD in transmales with a bone age of <14 years was statistically significantly higher at 24-month follow-up compared with start of gender-affirming hormones (z-score [range]: start of hormones -0.84 [-2.2 to 0.87], 24-month follow-up -0.15 [-1.38 to 0.94], p≤ 0.01). Statistically significant improvements in z-score for lumbar spine BMAD in transmales with a bone age of ≥14 years were also seen (z-score [range]: start of hormones -0.29 [-2.28 to 0.90], 24-month follow-up -0.06 [-1.75 to 1.61], p≤ 0.01).
- Actual lumbar spine BMAD values in g/cm^3 were statistically significantly higher at 24-month follow-up compared with start of gender-affirming hormones in transfemales and transmales of all bone ages **(VERY LOW)**.

Bone mineral density (BMD)

Two uncontrolled, observational studies reported change in lumbar BMD (Klink et al. 2015; Stoffers et al. 2019). BMD was determined using dual energy x-ray absorptiometry (DXA-scan; HologicQDR4500, Hologic). BMD was reported as g/cm^2 and as z-scores – see BMAD above for more details).

In Klink et al. 2015 (n=34):

- There was no statistically significant difference in lumbar spine BMD z-score from starting gender-affirming hormones to age 22 years in transfemales or transmales.
- Actual lumbar spine BMD values in g/cm^2 were statistically significantly higher at age 22 years compared with the start of gender-affirming hormones in transfemales and transmales **(VERY LOW)**.

In Stoffers et al. 2019 (n=62 at 6-month follow-up; n=15 at 24-month follow-up):

- There was no statistically significant difference in lumbar spine BMD z-score in transmales from starting gender-affirming hormones to any timepoint (6, 12 and 24 months).
- There was also no statistically significant difference in actual lumbar spine BMD values in g/cm^2 from starting gender-affirming hormones to any timepoint (6, 12 and 24 months) **(VERY LOW)**.

	These studies provide very low certainty evidence that lumber spine bone density (measured by BMAD) increases during treatment with gender-affirming hormones (from baseline to follow-up of 2 to 5 years). Z-scores at the end of follow-up suggest the average lumbar spine bone density was generally lower than the equivalent cisgender population (transfemales compared with cis-males and transmales compared with cis-females). The results for bone density (measured by BMD) were inconsistent.
Change in bone density: femoral neck **Certainty of evidence: very low**	This is an important outcome because childhood and adolescence is a key time for bone development and gender-affirming hormones may affect bone development, as shown by changes in femoral neck bone density. Three uncontrolled, observational studies (2 retrospective and 1 prospective) provided evidence related to bone density: femoral neck in children and adolescents with gender dysphoria. This was reported as either bone mineral density (BMD), bone mineral apparent density (BMAD), or both. One study reported change in bone density from start of gender-affirming hormones to age 22 years (Klink et al. 2015). Two studies reported change in bone density from start of gender-affirming hormones up to 24-month follow-up (Stoffers et al. 2019 and Vlot et al. 2017). All participants had previously been treated with a GnRH analogue. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. **Bone mineral apparent density (BMAD)** Two uncontrolled, observational studies reported change in femoral neck BMAD (Klink et al. 2015; Vlot et al. 2017). See above for more details on BMAD. In Klink et al. 2015 (n=34): • The z-score for femoral neck BMAD was reported for the start of gender-affirming hormones but not at age 22 years in transfemales or transmales. No statistical analysis reported. • In transfemales there was no statistically significant difference in actual femoral neck BMAD values in g/cm^3 at age 22 years compared with start of gender-affirming hormones. In transmales actual lumbar spine BMAD values in g/cm^3 were statistically significantly higher at age 22 years compared with start of gender-affirming hormones (mean [±SD]: start of hormones 0.31 [±0.04], age 22 years 0.33 [±0.05], p=0.010) **(VERY LOW)**. In Vlot et al. 2017 (n=70): • In transfemales (all bone ages), there was no statistically significant difference in femoral neck BMAD z-score from start of gender-affirming hormones to 24-month follow-up. • The z-score for femoral neck BMAD in transmales with a bone age of <14 years was statistically significantly higher at 24-month follow-up compared with start of gender-affirming hormones (z-score [range]: start of hormones -0.37 [-2.28 to 0.47], 24-month follow-up -0.37 [-2.03 to 0.85], p≤0.01). Statistically significant improvements in z-score for lumbar spine BMAD in transmales with a bone age of ≥14 years were also

	seen (z-score [range]: start of hormones -0.27 [-1.91 to 1.29], 24-month follow-up 0.02 [-2.1 to 1.35], p≤0.05). • In transfemales of all bone ages, there was no statistically significant change in actual femoral neck BMAD values in g/cm^3 from start of gender-affirming hormones to 24-month follow-up. In transmales of all bone ages, actual femoral neck BMAD values in g/cm^3 were statistically significantly higher at 24-month follow-up compared with start of gender-affirming hormones **(VERY LOW)**. **Bone mineral density (BMD)** Two uncontrolled, observational studies reported change in femoral neck BMD (Klink et al. 2015; Stoffers et al. 2019). See above for more details on BMD. In Klink et al. 2015 (n=34): • In transfemales, there was no statistically significant difference in femoral neck BMD z-score from start of gender-affirming hormones to age 22 years. In transmales, femoral neck BMD z-score was statistically significantly higher at age 22 years compared with start of gender-affirming hormones (z-score [SD]: start of hormones -0.35 [0.79], age 22 years -0.35 [0.74], p=0.006). • Actual femoral neck BMD values in g/cm^2 were statistically significantly higher at age 22 years compared with start of gender-affirming hormones in transfemales and transmales **(VERY LOW)**. In Stoffers et al. 2019 (n=62 at 6-month follow-up; n=15 at 24-month follow-up): • there was no statistically significant difference in right or left femoral neck BMD z-score in transmales, from the start of gender-affirming hormones to any timepoint (6, 12 and 24 months). • There was also no statistically significant difference in transmales in right or left actual femoral neck BMD values in g/cm^2 from start of gender-affirming hormones to any timepoint (6, 12 and 24 months) **(VERY LOW)**. **These studies provide very low certainty evidence that during treatment with gender-affirming hormones from baseline to follow-up of 2 to 5 years, femoral neck bone density (measured by BMAD) was unchanged in transfemales but was statistically significantly increased in transmales (although the absolute change was small). Z-scores at the end of follow-up suggest that average femoral neck bone density was lower in both transfemales and transmales than in the equivalent cisgender population (transfemales compared with cis-males and transmales compared with cis-females). The results for bone density (measured by BMD) were inconsistent.**
Change in clinical parameters: glucose, insulin and HbA1c	This is an important outcome because the effect of gender-affirming hormones on insulin sensitivity and cardiovascular risk in children and adolescents with gender dysphoria is unknown.

Certainty of evidence: very low	Two uncontrolled, retrospective chart reviews (Klaver et al. 2020; Stoffers et al. 2019) provided evidence on glucose, insulin and HbA1c. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. **Glucose levels, insulin levels and insulin resistance** One retrospective chart review (Klaver et al. 2020) reported non-comparative evidence on the change in glucose levels, insulin levels and insulin resistance (measured using Homeostatic Model Assessment of Insulin Resistance [HOMA-IR]) between starting gender-affirming hormones and age 22 years. In Klaver et al. 2020 (n=192): • There was no statistically significant change in glucose levels, insulin levels and insulin resistance in transfemales. • There was no statistically significant change in glucose levels in transmales. • There was a statistically significant decrease in insulin levels in transmales (mean change [95% CI] -2.1 mU/L [-3.9 to -0.3], p<0.05; mean insulin level at 22 years [95% CI] 8.6 mU/L [6.9 to 10.2]). • There was a statistically significant decrease in insulin resistance in transmales (HOMA-IR; mean change [95% CI] -0.5 [-1.0 to -0.1], p<0.05; mean HOMA-IR at 22 years [95% CI] 1.8 [1.4 to 2.2]) **(VERY LOW)**. **HbA1c** One retrospective chart review (Stoffers et al. 2019; n=62) reported non-comparative evidence on the change in HbA1c in transmales between starting gender-affirming hormones and 24-month follow-up. There was no statistically significant change in HbA1c **(VERY LOW)**. **These studies provide very low certainty evidence that gender-affirming hormones do not affect HbA1c, glucose levels, insulin levels and insulin resistance.**
Change in clinical parameters: lipids **Certainty of evidence: very low**	This is an important outcome because the effect of gender-affirming hormones on lipid profiles and cardiovascular risk in children and adolescents with gender dysphoria is unknown. One retrospective chart review (Klaver et al. 2020) provided non-comparative evidence on the change in lipids (total cholesterol, HDL cholesterol, LDL cholesterol and triglycerides) between starting gender-affirming hormones and age 22 years. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. In Klaver et al. 2020 (n=192): • There was no statistically significant change in total cholesterol, HDL cholesterol and LDL cholesterol in transfemales. • There was a statistically significant decrease (improvement) in triglycerides in transfemales (mean change [95% CI] +0.2 mmol/L [0.0 to 0.5], p<0.05; mean triglyceride level at 22 years [95% CI] 1.1 mmol/L [0.9 to 1.4]). • There was a statistically significant increase in total cholesterol in transmales (mean change [95% CI] +0.4 mmol/L [0.2 to 0.6],

	p<0.001; mean total cholesterol at 22 years [95% CI] 4.6 mmol/L [4.3 to 4.8]). • There was a statistically significant decrease (worsening) in HDL cholesterol (mean change in transmales [95% CI] -0.3 mmol/L [-0.4 to -0.1], p<0.001; mean HDL cholesterol at 22 years [95% CI] 1.3 mmol/L [1.2 to 1.3]). • There was a statistically significant increase (worsening) in LDL cholesterol in transmales (mean change [95% CI] +0.4 mmol/L [0.2 to 0.6], p<0.001; mean LDL cholesterol at 22 years [95% CI] 2.6 mmol/L [2.4 to 2.8]). • There was a statistically significant increase (worsening) in triglycerides in transmales (mean change [95% CI] +0.5 mmol/L [0.3 to 0.7], p<0.001; mean triglyceride level at 22 years [95% CI] 1.3 mmol/L [1.1 to 1.5]) **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones do not affect lipid profiles in transfemales. In transmales, there was a small but statistically significant worsening in cholesterol levels from start of gender-affirming hormone treatment to age 22 years, but mean cholesterol and triglyceride levels were within the UK reference range at the end of treatment.**
Change in clinical parameters: blood pressure **Certainty of evidence: very low**	This is an important outcome because the effect of gender-affirming hormones on blood pressure and cardiovascular risk in children and adolescents with gender dysphoria is unknown. One retrospective chart review (Klaver et al. 2020) provided non-comparative evidence on the change in blood pressure between starting gender-affirming hormones and at age 22 years. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. In Klaver et al. 2020 (n=192): • There was no statistically significant change in systolic blood pressure (SBP) in transfemales. However, there was a statistically significant increase in diastolic blood pressure (DBP) in transfemales (mean change [95% CI] +6 mmHg [3 to 10], p<0.001; mean DBP at 22 years [95% CI] 75 [72 to 78]). • In transmales, there was a statistically significant increase in SBP (mean change [95% CI] +5 mmHg [1 to 9], p<0.05; mean SBP at 22 years [95% CI] 126 [122 to 130]), and DBP (mean change [95% CI] +6 mmHg [4 to 9], p<0.001; mean DBP at 22 years [95% CI] 74 [72 to 77]) **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase blood pressure from start of treatment to age 22 years, although the absolute increase was small.**
Change in clinical parameters: body mass index (BMI)	This is an important outcome because the effect of gender-affirming hormones on weight gain and cardiovascular risk in children and adolescents with gender dysphoria is unknown. One retrospective chart review (Klaver et al. 2020) provided non-comparative evidence on the change in body mass index (BMI) between starting gender-affirming hormones and age 22 years. All

Certainty of evidence: very low	outcomes were reported separately for transfemales and transmales; also see subgroups table below. In Klaver et al. 2020 (n=192): • There was a statistically significant increase in BMI in transfemales from the start of gender-affirming hormones to age 22 years (mean change [95% CI] +1.9 [0.6 to 3.2], p<0.005; mean BMI at 22 years [95% CI] 23.2 [21.6 to 24.8]. At age 22 years, 9.9% of transfemales were obese, compared with 3.0% in a reference population of cisgender men. • There was a statistically significant increase in BMI in transmales from the start of gender-affirming hormones to age 22 years (mean change [95% CI] +1.4 [0.8 to 2.0], p<0.005; mean BMI at 22 years [95% CI] 23.9 [23.0 to 24.7]). At age 22 years, 6.6% of transmales were obese, compared with 2.2% in a reference population of cisgender women **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase BMI from start of treatment to age 22 years, although most participants were within the healthy weight range.**
Change in clinical parameters: liver function **Certainty of evidence: very low**	This is an important outcome because if treatment-induced liver injury (raised liver enzymes are a marker of this) is suspected, gender-affirming hormones may need to be stopped. One retrospective chart review (Stoffers et al. 2019) provided non-comparative evidence on the change in liver enzymes in transmales between starting gender-affirming hormones and up to 24-months follow-up. In Stoffers et al. 2019 (n=62): • There was no statistically significant change in aspartate aminotransferase (AST), alanine aminotransferase (ALT) and gamma-glutamyltransferase (GCT) in transmales. • There was a statistically significant increase in alkaline phosphatase (ALP) levels from starting gender-affirming hormones to 6- and 12-months follow-up, although by 24-months the difference was not statistically significant (median [IQR]: start of hormones 102 [78 to 136], 6-month follow-up 115 [102 to 147] p<0.001, 12-month follow-up 112 [88 to 143] p<0.001) **(VERY LOW)**. **This study provides very low certainty evidence that gender-affirming hormones do not affect liver function in transmales from baseline to 24 months follow-up.**
Change in clinical parameters: kidney function **Certainty of evidence: very low**	This is an important outcome because if renal damage (raised serum creatinine and urea are markers of this) is suspected, treatment with gender-affirming hormones may need to be stopped. One retrospective chart review (Stoffers et al. 2019) provided non-comparative evidence on the change in serum creatinine and serum urea levels in transmales between starting gender-affirming hormones and up to 24-months follow-up. In Stoffers et al. 2019 (n=62):

	• There was a statistically significant increase in creatinine levels in transmales at all timepoints up to 24 months (mean [SD]: start of hormones 62 umol/L [7], 6 months 70 umol/L [9] , 12 months 74 umol/L [10], 24 months 81 umol/L [10], p<0.001). • There was no statistically significant change in urea in transmales (follow-up duration not reported) **(VERY LOW)**. **This study provides very low certainty evidence on the effects of gender-affirming hormones on kidney function in transmales from baseline to 24 months follow-up. A statistically significant increase in creatinine levels was seen, but these were within the UK reference range. Urea levels were unchanged.**

Treatment discontinuation
 	
Certainty of evidence: very low
	This is an important outcome because there is uncertainty about the short- and long-term impact of stopping treatment with gender-affirming hormones in children and adolescents with gender dysphoria.
	
	One uncontrolled, retrospective chart review (Khatchadourian et al. 2014) provided evidence relating to permanent or temporary treatment discontinuation in children and adolescents with gender dysphoria.
	
	Khatchadourian et al. 2014 narratively reported treatment discontinuation in a cohort of 63 adolescents (24 transfemales and 39 transmales) who received gender-affirming hormones:
	• No participants permanently discontinued gender-affirming hormones.
	• No transfemales temporarily discontinued gender-affirming hormones.
	• Three transmales temporarily discontinued gender-affirming hormones due to:
	  ○ mental health comorbidities (n=2)
	  ○ androgenic alopecia (n=1).
	All 3 participants eventually resumed treatment, although timescales were not reported **(VERY LOW)**.
	
	This study provides very low certainty evidence that the rates of discontinuation during treatment with gender-affirming hormones are low (duration of treatment not reported).

Adverse effects
 	
Certainty of evidence: very low
	This is an important outcome because if there are adverse effects, gender-affirming hormones may need to be stopped.
	
	One uncontrolled, retrospective chart review (Khatchadourian et al. 2014) provided evidence relating to adverse effects from gender-affirming hormones in children and adolescents with gender dysphoria.
	
	Khatchadourian et al. 2014 narratively reported adverse effects in a cohort of 63 adolescents (24 transfemales and 39 transmales) receiving treatment with gender-affirming hormones:
	• No severe complications were reported.
	• No transfemales reported minor complications.
	• Twelve transmales developed minor complications, which were:
	  ○ severe acne, requiring isotretinoin treatment (n=7)
	  ○ androgenic alopecia (n=1)
	  ○ mild dyslipidaemia (further details not provided; n=3)
	  ○ significant mood swings (n=1) **(VERY LOW)**.

	This study provides very low certainty evidence about the potential adverse effects of gender-affirming hormones (duration of treatment not reported). No conclusions could be drawn.

Abbreviations: ALP: alkaline phosphatase; ALT: alanine aminotransferase; AST: aspartate aminotransferase; BMAD: bone mineral apparent density; BMD: bone mineral density; BMI: body mass index; DBP: diastolic blood pressure; GGT: gamma-glutamyl transferase; HbA1c: glycated haemoglobin; HDL: high-density lipoproteins; HOMA-IR: Homeostatic Model Assessment of Insulin Resistance; IQR: interquartile range; LDL: low-density lipoproteins; p: p-value; SBP: systolic blood pressure; SD: standard deviation.

In children and adolescents with gender dysphoria, what is the cost-effectiveness of gender-affirming hormones compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement
Cost-effectiveness	No studies were identified to assess the cost-effectiveness of gender-affirming hormones for children and adolescents with gender dysphoria.

From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may benefit from gender-affirming hormones more than the wider population of interest?

Subgroup	Evidence statement
Sex assigned at birth males (transfemales) **Certainty of evidence: Very low**	Some studies reported data separately for sex assigned at birth males (transfemales). This included some direct comparisons with sex assigned at birth females (transmales). **Impact on mental health: depression and anxiety** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported the change in depression (measured using QIDS clinician-reported and self-reported), anxiety and anxiety-related symptoms (measured using SCARED) in transfemales. See the clinical effectiveness results above for full details. In Kuper et al. 2020 (n=33 to 45, varies by outcome), changes were seen in depression, anxiety and anxiety-related symptoms from baseline to follow-up but the authors did not report any statistical analyses, so it is unclear if was any changes were statistically significant **(VERY LOW)**. **This study provides very low certainty evidence on the effects of gender-affirming hormones on depression, anxiety and anxiety-related symptoms over time in sex assigned at birth males (transfemales; mean duration of treatment 10.9 months). No conclusions could be drawn.** **Impact on mental health: suicidality**

	One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in Ask Suicide-Screening Questions (ASQ) in transfemales compared with transmales. See the clinical effectiveness results above for full details. Between baseline and the final assessment, there was no statistically significant difference in change in ASQ score for transfemales compared with transmales (p=0.79; n=47) **(VERY LOW)**. One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in suicidal ideation in transfemales measured using additional questions from the PHQ 9_Modified for Teens. See the clinical effectiveness results above for full details. At baseline, 11.8% (2/17) of transfemales had suicidal ideation, compared with 5.9% (1/17) at about 12-months follow-up (no statistical analysis reported) **(VERY LOW)**. **These studies provide very low certainty evidence that any change in suicidal ideation is not different between sex assigned at birth males (transfemales) and sex assigned at birth females (transmales) from baseline to follow-up of about 12 months.** **Impact on quality of life** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in the GWBS of the Paediatric Quality of Life Inventory in transfemales compared with transmales. See the clinical effectiveness results above for full details. Between baseline and final assessment, there was no statistically significant difference in change in GWBS of the Paediatric Quality of Life Inventory for transfemales compared with transmales (p=0.32; n=47) **(VERY LOW)**. **This study provides very low certainty evidence that any change in general wellbeing is not different between sex assigned at birth males (transfemales) and sex assigned at birth females (transmales) from baseline to follow-up of about 12 months.** **Impact on body image** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported change in Body Image Scale (BIS) in transfemales. See the clinical effectiveness results above for full details. In Kuper et al. 2020 (n=30), the mean (±SD) BIS score was 67.5 points (±19.5) at baseline and 49.0 points (±21.6) at follow-up (no statistical analysis reported) **(VERY LOW).** **This study provides very low certainty evidence on the effects of gender-affirming hormones on body image over time in transfemales (mean duration of treatment 10.9 months). No conclusions could be drawn.** **Change in bone density: lumbar spine**

	Two uncontrolled, observational, retrospective studies provided evidence relating to the effect of gender-affirming hormones on lumber spine bone density in transfemales (Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results. **These studies provide very low certainty evidence that lumbar spine bone density (measured by BMAD) increases during treatment with gender-affirming hormones in sex assigned at birth males (transfemales). Z-scores at the end of follow-up suggest average lumbar spine bone density was generally lower than in the equivalent cisgender population. The results for lumbar spine bone density (measured by BMD) were inconsistent.** **Change in bone density: femoral neck** Two uncontrolled, observational, retrospective studies provided evidence relating to the effect of gender-affirming hormones on femoral neck bone density in transfemales (Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results. **These studies provide very low certainty evidence that femoral neck bone density (measured by BMAD) was unchanged in sex assigned at birth males (transfemales) during treatment with gender-affirming hormones (follow-up between 2 and 5 years). Z-scores at the end of follow-up suggest and the average femoral neck bone density was lower than in the equivalent cisgender population. The results for femoral neck bone density (measured by BMD) were inconsistent.** **Change in clinical parameters: glucose, insulin and HbA1c** One uncontrolled, retrospective chart review (Klaver et al. 2020) provided evidence on glucose, insulin and HbA1c in transfemales. See the safety results table above for a full description of the results. **This study provided very low certainty evidence that gender-affirming hormones do not affect HbA1c, glucose levels, insulin levels and insulin resistance in sex assigned at birth males (transfemales) from the start of treatment to age 22 years.** **Change in clinical parameters: lipids** One retrospective chart review (Klaver et al. 2020) provided evidence on the change in lipids (total cholesterol, HDL cholesterol, LDL cholesterol and triglycerides) in transfemales. See the safety results table above for a full description of the results. **This study provides very low certainty evidence that gender-affirming hormones do not affect lipid profiles in sex assigned at birth males (transfemales) from the start of treatment to age 22 years.** **Change in clinical parameters: blood pressure**

	One retrospective chart review (Klaver et al. 2020) provided evidence on the change in blood pressure in transfemales. See the safety results table above for a full description of the results. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase blood pressure in sex assigned at birth males (transfemales), although the absolute increase was small from the start of treatment to age 22 years.** **Change in clinical parameters: body mass index (BMI)** One retrospective chart review (Klaver et al. 2020) provided evidence on the change in BMI in transfemales. See the safety results table above for a full description of the results. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase BMI in sex assigned at birth males (transfemales), although most participants were within the healthy weight range from the start of treatment to age 22 years.** **Treatment discontinuation** One uncontrolled, retrospective chart review provided evidence relating to permanent or temporary discontinuation of gender-affirming hormones in transfemales (Khatchadourian et al. 2014). **This study provides very low certainty evidence that the rates of discontinuation during treatment with gender-affirming hormones in sex assigned at birth males (transfemales) are low. Duration of treatment with gender-affirming hormones was not reported.** **Adverse effects** One uncontrolled, retrospective chart review provided evidence relating to adverse effects from gender-affirming hormones in transfemales (Khatchadourian et al. 2014). **This study provides very low certainty evidence about the potential adverse effects of gender-affirming hormones in sex assigned at birth males (transfemales). No conclusions could be drawn. Duration of treatment with gender-affirming hormones was not reported.**
Sex assigned at birth females (transmales) **Certainty of evidence: Very low**	Some studies reported data separately for sex assigned at birth females (transmales). This included some direct comparisons with sex assigned at birth males (transfemales). **Impact on mental health: depression and anxiety** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported the change in depression (measured using QIDS clinician-reported and self-reported), anxiety and anxiety-related symptoms (measured using SCARED) in transmales. See the clinical effectiveness results above for full details. In Kuper et al. 2020 (n=65 to 78, varies by outcome), changes were seen in depression, anxiety and anxiety-related symptoms from

	baseline to follow-up but the authors did not report any statistical analysis, so it is unclear if any changes are statistically significant **(VERY LOW)**. **This study provides very low certainty evidence on the effects of gender-affirming hormones on depression, anxiety and anxiety-related symptoms over 10.9 months in transmales. No conclusions could be drawn.** **Impact on mental health: suicidality** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in Ask Suicide-Screening Questions (ASQ) in transmales compared with transfemales. See the sex assigned at birth males (transfemales) row above for full details of the results. One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported the change in suicidal ideation in transmales measured using additional questions from the PHQ 9_Modified for Teens. See the clinical effectiveness results above for full details. At baseline, 9.1% (3/33) of transmales had suicidal ideation, compared with 6.1% (2/33) at about 12-months follow-up (no statistical analysis reported) **(VERY LOW)**. **These studies provide very low certainty evidence that any change in suicidal ideation is not different between sex assigned at birth females (transmales) and sex assigned at birth males (transfemales). Mean duration of treatment about 12 months.** **Impact on quality of life** One uncontrolled, retrospective, longitudinal study (Allen et al. 2019) reported the change in the GWBS of the Paediatric Quality of Life Inventory in transmales compared with transfemales. See the sex assigned at birth males (transfemales) row above for full details of the results. **This study provides very low certainty evidence that any change in general wellbeing is not different between sex assigned at birth females (transmales) and sex assigned at birth males (transfemales). Mean duration of treatment about 12 months.** **Impact on body image** One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported change in Body Image Scale (BIS) in transmales. See the clinical effectiveness results above for full details. In Kuper et al. 2020 (n=66), the mean (±SD) BIS score was 71.1 points (±13.4) at baseline and 52.9 points (±16.8) at follow-up (no statistical analysis reported) **(VERY LOW)**. **This study provides very low certainty evidence on the effects of gender-affirming hormones on body image over 10.9 months in transmales. No conclusions could be drawn.** **Change in bone density: lumbar spine**

	Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of gender-affirming hormones on lumber spine bone density in transmales (Klink et al. 2015, Stoffers et al. 2019 and Vlot et al. 2017). See the safety results table above for a full details of the results. **These studies provide very low certainty evidence that lumbar spine bone density (measured by BMAD) increases during 2 to 5 years treatment with gender-affirming hormones in sex assigned at birth females (transmales). Z-scores at the end of follow-up suggest the average lumbar spine bone density was generally lower than in the equivalent cisgender population. The results for lumbar spine bone density (measured by BMD) were inconsistent.** **Change in bone density: femoral neck** Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of gender-affirming hormones on femoral neck bone density in transmales (Klink et al. 2015, Stoffers et al. 2019 and Vlot et al. 2017). See the safety results table above for a full details of the results. **These studies provide very low certainty evidence that femoral neck bone density (measured by BMAD) statistically significantly increased in sex assigned at birth females (transmales) during 2 to 5 years treatment with gender-affirming hormones. Z-scores at the end of follow-up suggest the average femoral neck bone density was generally lower than in the equivalent cisgender population. The results for femoral neck bone density (measured by BMD) were inconsistent.** **Change in clinical parameters: glucose, insulin and HbA1c** Two uncontrolled, retrospective chart reviews (Klaver et al. 2020; Stoffers et al. 2019) provided evidence on glucose, insulin and HbA1c in transmales. See the safety results table above for full details of the results. **This study provided very low certainty evidence that gender-affirming hormones do not affect HbA1c, glucose levels, insulin levels and insulin resistance in sex assigned at birth females (transmales). Reported from start of treatment to age 22 years.** **Change in clinical parameters: lipids** One retrospective chart review (Klaver et al. 2020) provided evidence on the change in lipids (total cholesterol, HDL cholesterol, LDL cholesterol and triglycerides) in transmales. See the safety results table above for full details of the results. **This study provides very low certainty evidence that treatment with gender-affirming hormones is associated with a small but statistically significant worsening of cholesterol levels in sex assigned at birth females (transmales), but mean cholesterol and triglyceride levels were within the UK reference range at end of treatment, from start of treatment to age 22 years.**

	Change in clinical parameters: blood pressure One retrospective chart review (Klaver et al. 2020) provided evidence on the change in blood pressure in transmales. See the safety results table above for full details of the results. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase blood pressure in sex assigned at birth females (transmales), although the absolute increase was small, from start of treatment to age 22 years.** **Change in clinical parameters: body mass index (BMI)** One retrospective chart review (Klaver et al. 2020) provided evidence on the change in body mass index (BMI) in transmales. See the safety results table above for full details of the results. **This study provides very low certainty evidence that gender-affirming hormones statistically significantly increase BMI in sex assigned at birth females (transmales), although most participants were within the healthy weight range, from start of treatment to age 22 years.** **Change in clinical parameters: liver function** One retrospective chart review (Stoffers et al. 2019) provided non-comparative evidence on the change in liver enzymes in transmales between starting gender-affirming hormones and up to 24-months follow-up. See the safety results table above for full details of the results. **This study provides very low certainty evidence that gender-affirming hormones for about 12 months do not affect liver function in sex assigned at birth females (transmales).** **Change in clinical parameters: kidney function** One retrospective chart review (Stoffers et al. 2019) provided non-comparative evidence on the change in serum creatinine and serum urea levels in transmales between starting gender-affirming hormones and up to 24-months follow-up. See the safety results table above for full details of the results. **This study provides very low certainty evidence on the effects of gender-affirming hormones on kidney function in sex assigned at birth females (transmales). A statistically significant increase in creatinine levels was seen at about 12 months follow-up, but these were within the UK reference range. Urea levels were unchanged.** **Treatment discontinuation** One uncontrolled, retrospective chart review provided evidence relating to permanent or temporary discontinuation of gender-affirming hormones in transmales (Khatchadourian et al. 2014). See the safety results table above for full details of the results. **This study provides very low certainty evidence that the rates of treatment discontinuation with gender-affirming hormones in sex**

	assigned at birth females (transmales) is low. Duration of gender-affirming hormones not reported. **Adverse effects** One uncontrolled, retrospective chart review provided evidence for adverse effects from gender-affirming hormones in transmales (Khatchadourian et al. 2014). See the safety results table above for full details of the results. **This study provides very low certainty evidence about the potential adverse effects of gender-affirming hormones in sex assigned at birth females (transmales). No conclusions could be drawn. Duration of gender-affirming hormones not reported.**
Duration of gender dysphoria	No evidence was identified.
Age at onset of gender dysphoria	No evidence was identified.
Age at onset of puberty	No evidence was identified.
Tanner stage at which GnRH analogue or gender-affirming hormones started	One uncontrolled, prospective, longitudinal study (Kuper et al. 2020) reported the impact of Tanner stage on outcomes, although it is not clear whether this is referring to Tanner stage at initial assessment, at the start of GnRH analogues or at another timepoint.
Diagnosis of autistic spectrum disorder	No evidence was identified.
Diagnosis of a mental health condition	One uncontrolled, prospective, longitudinal study (Achille et al. 2020) reported outcomes that were adjusted for engagement in counselling and medicines for mental health problems. Information about diagnoses and treatment were not provided. Rates of mental health issues appear to be high in the cohort. **Impact on mental health** Achille et al. 2020 reported the change in depression scores, controlled for engagement in counselling and medicines for mental health problems (measured using the Center for Epidemiologic Studies Depression [CESD-R] scale and Patient Health Questionnaire Modified for Teens [PHQ 9_Modified for Teens] score: • There was no statistically significant change in CESD-R from baseline to about 12-months follow-up. • There was no statistically significant change in PHQ 9_Modified for Teens score from baseline to about 12-months follow-up **(VERY LOW)**. **Impact on quality of life** Achille et al. 2020 reported the change in quality of life scores, controlled for engagement in counselling and medicines for mental health problems (measured using the Quality of Life Enjoyment and Satisfaction Questionnaire [QLES-Q-SF] score: • There was no statistically significant change in QLES-Q-SF score from baseline to about 12-months follow-up **(VERY LOW)**.

	This study provides very low certainty evidence about outcomes that were adjusted for engagement in counselling and medicines for mental health problems. No conclusions could be drawn.

Abbreviations: ASQ: Ask Suicide-Screening Questions; CESD-R: Center for Epidemiologic Studies Depression; GnRH: Gonadotrophin releasing hormone; GWBS: General Well-Being Scale; HDL: high-density lipoproteins; LDL: low-density lipoproteins; p: p-value; PHQ 9_Modified for Teens: Patient Health Questionnaire Modified for Teens; QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire.

From the evidence selected,

(a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?

(b) what were the ages at which participants commenced treatment with gender-affirming hormones?

(c) what was the duration of treatment with GnRH analogues?

<table>
<tr><th>Outcome</th><th>Evidence statement</th></tr>
<tr><td>Diagnostic criteria</td><td>The DSM-IV-TR criteria was used in 3 studies (Klaver et al. 2020, Klink et al. 2015 and Vlot et al. 2017).

The DSM-V criteria was used in 2 studies (Kuper et al. 2020 and Stoffers et al. 2019). The DSM-V has one overarching definition of gender dysphoria with separate specific criteria for children and for adolescents and adults. The general definition describes a conflict associated with significant distress and/or problems functioning associated with this conflict between the way they feel and think of themselves which must have lasted at least 6 months.

The ICD-10 diagnosis of ‘transsexualism’ was used in 1 study (Kaltiala et al. 2020). The authors state that this is the corresponding diagnosis to ‘gender dysphoria’ in the DSM-V, and that diagnostic assessments in the study location (Finland) take place according to ICD-10.

It was not reported how gender dysphoria was defined in the remaining 4 studies (VERY LOW).

From the evidence selected, the most commonly reported diagnostic criteria for gender dysphoria (5/10 studies) was the DSM criteria in use at the time the study was conducted.</td></tr>
<tr><td>Age when gender-affirming hormones started</td><td>8/10 studies reported the age at which participants started treatment with gender-affirming hormones, either as the mean age (with SD) or median age (with the range):

<table>
<tr><th>Study</th><th>Mean age (± SD)</th></tr>
<tr><td>Allen et al. 2019</td><td>16.7 years (not reported)</td></tr>
<tr><td>Khatchadourian et al. 2014</td><td>17.4 years (1.9)</td></tr>
<tr><td>Klaver et al. 2020</td><td>16.4 years (1.1) in transfemales
16.9 years (0.9) in transmales</td></tr>
<tr><td>Kuper et al. 2020</td><td>16.2 (1.2)</td></tr>
<tr><td>Klink et al. 2015</td><td>16.6 years (1.4) in transfemales
16.4 years (2.3) in transmales</td></tr>
</table>
</td></tr>
</table>

<table>
<tr><td></td><td>
<table>
<tr><th>Study</th><th>Median age (range)</th></tr>
<tr><td>Stoffers et al. 2019</td><td>17.2 years (15 to 19.5)</td></tr>
<tr><td>Vlot et al. 2017</td><td>16.3 years (15.9 to 19.5) in transfemales
16.0 years (14.0 to 18.9) in transmales</td></tr>
</table>

Age at the start of treatment was not reported in 3 studies:
• In Achille et al. 2020 the mean age at initial assessment (baseline) was 16.2 years (SD ±2.2)
• In Kaltiala et al. 2020 the mean age at diagnosis was 18.1 years (range 15.2 to 19.9)
• In Lopez de Lara et al. 2020 the mean age of participants was 16 years (range 14 to 18), although it is not clear if this is at the initial assessment or at the start of gender-affirming hormones.

The evidence included showed that most children and adolescents started treatment with gender-affirming hormones at about 16 to 17 years, with a range of about 14 to 19 years.
</td></tr>
<tr><td>Duration of treatment with GnRH analogues</td><td>
The duration of treatment with GnRH analogues was reported in 3/10 studies:
<table>
<tr><th>Study</th><th>Median duration</th></tr>
<tr><td>Klaver et al. 2020</td><td>2.1 years (IQR 1.0 to 2.7) in transfemales
1.0 years (IQR 0.5 to 2.9) in transmales</td></tr>
<tr><td>Klink et al. 2015</td><td>1.3 years (range 0.5 to 3.8) in transfemales
1.5 years (range 0.25 to 5.2) in transmales
(GnRH analogue monotherapy)</td></tr>
<tr><td>Stoffers et al. 2019</td><td>8 months (range 3 to 39)</td></tr>
</table>

The evidence included showed wide variation in the duration of treatment with gender-affirming hormones, but most studies did not report this information. Treatment duration ranged from a few months up to about 5 years.
</td></tr>
</table>

Abbreviations: DSM, Diagnostic and Statistical Manual of Mental Disorders criteria; GnRH, Gonadotrophin-releasing hormone; ICD, International Statistical Classification of Diseases and Related Health Problems; IQR, interquartile range; SD, standard deviation.

6. Discussion

A key limitation to identifying the effectiveness and safety of gender-affirming hormones for children and adolescents with gender dysphoria is the lack of reliable comparative studies. All the studies included in this evidence review are uncontrolled observational studies, which are subject to bias and confounding and were of very low certainty using modified GRADE. The size of the population with gender dysphoria means conducting a prospective trial may be unrealistic, at least on a single centre basis. There may also be ethical issues with a 'no treatment arm' in comparative trials of gender-affirming hormones, where there may be poor mental health outcomes if treatment is withheld. However, the use of an active comparator such as close psychological support may reduce ethical concerns in future trials. A fundamental limitation of all the uncontrolled studies included in this review is that any changes in scores from baseline to follow-up could be attributed to a regression-to-the-mean.

The included studies have relatively short follow-up, with an average duration of treatment with gender-affirming hormones between around 1 year and 5.8 years. Further studies with a longer follow-up are needed to determine the long-term effect of gender-affirming hormones for children and adolescents with gender dysphoria.

Most studies included in this review did not report comorbidities (physical or mental health) and no study reported concomitant treatments in detail. Because of this it is not clear whether any changes observed were due to gender-affirming hormones or other treatments the participants may have received. For example, we do not know if any improvement in depression symptom score over time was the result of gender-affirming hormones or the mental health support the person may be receiving (including medicines or counselling). This may be of particular importance for the mental health outcomes discussed in this review, since depression, anxiety and other related symptoms are common in children and adolescents with gender dysphoria. In Achille et al. 2020, at baseline around one-third of participants were taking medicines for mental health problems and around two-thirds reported being depressed in the past year. In Kaltiala et al. 2020, half the participants needed mental health treatment during and before gender identity assessment, with the most common reasons for treatment being depression, anxiety and suicidality. Only 1 study reported outcomes adjusted for engagement in counselling and medicines for mental health problems (Achille et al. 2020). This study found that gender-affirming hormones had no significant impact on depression and quality of life when adjusted for mental health care, despite significant approvements reported for the unadjusted results. However, it is not possible to draw conclusions on the impact of concurrent mental health treatment on the effect of gender-affirming hormones based on this study alone. Details of the mental health care provided are not reported in the study and results are presented for transfemales and transmales separately, resulting in small patient numbers and possible underpowering.

In most of the included studies, details of the gender-affirming hormone treatment regimens are poorly reported, with limited information provided about the medicines, doses and routes of administration used. It is not clear whether the interventions used in the studies are reflective of current UK practice for children and adolescents with gender dysphoria. There is also the suggestion that the hormone dose used in 1 study may have been too low; the authors of Klink et al. 2015 suggest that the relatively low initial dose of oestrogen for transfemales may be the reason for the observed lack of effect on lumber spine bone density. Duration of treatment with a GnRH analogue is also poorly reported and is only stated in 3/10 studies.

There is a degree of indirectness in some studies, with some participants included that fall outside of the population of this evidence review. For example, in Kuper et al. 2020 17% of participants received puberty suppression alone, and in Achille et al. 2020, 30% of participants received no treatment or puberty suppression alone. Some results and statistical analyses are only reported for the whole cohort in these studies and not the subgroup of participants who received gender-affirming hormones.

Participant numbers are poorly reported in some of the included studies. In Achille et al. 2020, 47% (45/95) of the people who entered the study did not have follow-up data and were excluded from the analyses, with no explanation or description of those people lost to follow-up. In Kuper et al. 2020, the number of participants varied by outcome, with less than

two-thirds of participants providing data for some outcomes. The authors provide no explanation for this incomplete reporting.

It is not clear whether some outcome measures, specifically those related to psychosocial functioning, are relevant to the UK population. In Kaltiala et al. 2020, an observational study conducted in Finland, the proportion of participants living with parents or guardians is reported as marker of appropriate functioning. The authors state that in Finnish culture young people tend to leave the parental home early, with only around one-quarter of 20 to 24 year olds still living at home. This is lower than in the UK, where around half of 20 to 24 year olds live with their parents or guardians (ONS: Why are more young people living with their parents?).

It is difficult to draw firm conclusions for many of the effectiveness and safety outcomes reported in the included studies because many different scoring tools and methods were used to assess the same outcome, often with conflicting results. For example, bone density is reported as bone mineral density (BMD) and bone mineral apparent density (BMAD) in the same study, the latter being a size-adjusted measure often useful for people whose bones are still growing. For some populations (transfemale versus transmale) and bone regions (lumber spine versus femoral neck), statistically significant differences in BMD are reported but not for BMAD, and vice versa.

In addition to this, most outcomes reported across the included studies do not have an accepted minimal clinically important difference (MCID), making it difficult the determine whether any observed statistically significant changes are clinically meaningful. However, the authors of some studies report thresholds to interpret the results of the scoring tools, so some conclusions can be made. For example, the mean Utrecht Gender Dysphoria Scale (UGDS) score (a measure of gender dysphoria symptoms) reduced to about 15 points after treatment with gender-affirming hormones (Lopez de Lara et al. 2020). The authors state that scores of 40 points or above signify gender dysphoria, suggesting that after about 12 months of treatment with gender-affirming hormones, the majority of participants did not have symptoms of gender dysphoria.

The impact of gender-affirming hormones on bone density was reported in 3 studies (Klink et al. 2015, Stoffers et al. 2019 and Vlot et al. 2017). Although these studies did not include a control group, comparisons to a reference population are reported using z-scores. Comparisons were made to a cisgender population, meaning for example that bone density in transfemales was compared with bone density in cisgender males. The authors of Klink et al. 2015 note that this may not be the ideal comparison, because androgens and oestrogens affect bone differently, and that bone properties in a trans population differ from their age- and sex assigned at birth-matched controls. Beyond this, a major limitation when trying to determine the impact of gender-affirming hormones on the short- and long-term bone health of children and adolescents is the lack of data on fracture rates and other patient-orientated outcomes, including rates of osteoporosis. Studies of GnRH analogues in children and adolescents with gender dysphoria suggest that GnRH analogue treatment may reduce the expected increase in bone density (which is expected during puberty). Although improvements in bone density were reported following treatment with gender-affirming hormones, Z-scores suggest that bone density remained lower in transfemales and transmales compared with an equivalent cisgender population.

One study reported on cardiovascular risk factors at age 22 years in people who started gender-affirming hormones for gender dysphoria as adolescents. While glucose levels, insulin levels and insulin resistance were broadly unchanged at 22 years, statistically significant increases in blood pressure and body mass index were seen. A small but statistically significant worsening of the lipid profile in transmales who received testosterone was also seen at age 22 years. However, further studies with a considerably longer follow-up and a focus on patient-oriented outcomes, including cardiovascular events and mortality are needed to determine the long-term impact on cardiovascular health of starting gender-affirming hormones during childhood and adolescence.

Only 1 study reported adverse events and discontinuation rates with gender-affirming hormones in children and adolescents. Conclusions on these outcomes cannot be made based on this study alone.

This review did not identify sub-groups of people who may benefit more from gender-affirming hormones. Limited evidence from 2 studies suggests there was no difference in response to treatment between transfemales and transmales for mental health and quality of life (Achille et al. 2020 and Allen et al. 2019).

7. Conclusion

This evidence review found limited evidence for the effectiveness and safety of gender-affirming hormones in children and adolescents with gender dysphoria, with all studies being uncontrolled, observational studies, and all outcomes of very low certainty. Any potential benefits of treatment must be weighed against the largely unknown long-term safety profile of these treatments.

The results from 5 uncontrolled, observational studies (Achille et al. 2020, Allen et al. 2019, Kaltiala et al. 2020. Kuper et al. 2020, Lopez de Lara et al. 2020) suggest that, in children and adolescents with gender dysphoria, gender-affirming hormones are likely to improve symptoms of gender dysphoria, and may also improve depression, anxiety, quality of life, suicidality, and psychosocial functioning. The impact of treatment on body image is unclear. All results were of very low certainty. The clinical relevance of any improvements to the person is difficult to determine because most outcomes do not have a recognised minimal clinically important difference, and the authors do not present statistical analysis for some outcomes.

A further 5 uncontrolled, observational studies (Khatchadourian et al. 2014, Klaver et al. 2020, Klink et al. 2015, Stoffers et al. 2019 and Vlot et al. 2017) reported on safety outcomes, all of which provided very low certainty evidence. Statistically significant increases in some measures of bone density were seen following treatment with gender-affirming hormones, although results varied by bone region (lumber spine versus femoral neck) and by population (transfemales versus transmales). However, z-scores suggest that bone density remained lower in transfemales and transmales compared with an equivalent cisgender population. Results from 1 study of gender-affirming hormones started during adolescence reported statistically significant increases in blood pressure and body mass index, and worsening of the lipid profile (in transmales) at age 22 years, although longer term studies that report on cardiovascular event rates are needed. Adverse events and discontinuation rates associated with gender-affirming hormones were only reported in 1 study, and no conclusions can be made on these outcomes.

This review did not identify sub-groups of people who may benefit more from gender-affirming hormones. Limited evidence from 2 studies suggests there was no difference in response to treatment between transfemales and transmales for mental health and quality of life (Achille et al. 2020 and Allen et al. 2019).

No cost-effectiveness evidence was found to determine whether gender-affirming hormones are a cost-effective treatment for children and adolescents with gender dysphoria.

Appendix A PICO

The review questions for this evidence review are:

1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
2. For children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
3. For children and adolescents with gender dysphoria, what is the cost-effectiveness of gender-affirming hormones compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?
4. From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria?
5. From the evidence selected,
   (a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?
   (b) what were the ages at which participants commenced treatment with gender-affirming hormones?
   (c) what was the duration of GnRH analogues treatment?

PICO table

P –Population and Indication	Children and adolescents aged 18 years or less who have gender dysphoria, gender identity disorder or gender incongruence of childhood as defined by the study. The following subgroups of children and adolescents with gender dysphoria, gender identity disorder or gender incongruence of childhood need to be considered:

	• Sex assigned at birth males • Sex assigned at birth females • The duration of gender dysphoria: less than 6 months, 6-24 months, and more than 24 months) • The age at which treatment was initiated with GnRH analogues and with gender-affirming hormones. • The age of onset of gender dysphoria • The age of onset of puberty • Adolescents with gender dysphoria who have a pre-existing diagnosis of autistic spectrum disorder. • Adolescents with gender dysphoria who had a significant mental health symptom load at diagnosis including anxiety, depression (with or without a history of self-harm and suicidality), psychosis, personality disorder, Attention Deficit Hyperactivity Disorder and eating disorders.
I – Intervention	Gender-affirming hormone treatments: • A testosterone preparation for sex assigned at birth female patients which may include testosterone in the form of Sustanon injections*; testosterone enantate injections; Tostran gel*; Testogel; Testim gel; oral testosterone capsules in the form of testosterone undecanoate (Restandol); Andriol testocaps; Nebido • An oestradiol preparation** for sex assigned at birth male patients which may include: oral estradiol valerate*; oestrogen patches (7β-oestradiol patches e.g. Evorel or Estradem); Estradot patches; ethinyloestradiol *** *These are the used by Leeds Hospital, England. ** Be aware that the American spelling is oestrogen without the ‘o’. ***Ethinyloestradiol is rarely used.
C – Comparator(s)	One or a combination of: • Psychological support • Social transitioning to the gender with which the individual identifies. No intervention
O – Outcomes	There are no known minimal clinically important differences and there are no preferred timepoints for the outcome measures selected. **All outcomes should be stratified by:** • The age at which treatment with gender-affirming hormones was initiated • The length of treatment with GnRH analogues where possible. **A: Clinical Effectiveness** *Critical to decision making* • **Impact on gender dysphoria** This outcome is critical because gender dysphoria in adolescents and children is associated with significant distress and problems functioning. Impact on gender

	dysphoria may be measured by the Utrecht Gender Dysphoria Scale. Other measures as reported in studies may be used as an alternative to the stated measure. • **Impact on mental health** Examples of mental health problems include self-harm, thoughts of suicide, suicide attempts, suicide, eating disorders, depression/low mood and anxiety. These outcomes are critical because self-harm and thoughts of suicide have the potential to result in significant physical harm and for completed suicides the death of the young person. Disordered eating habits may cause significant morbidity in young people. Depression and anxiety are also critical outcomes because they may impact on social, occupational, or other areas of functioning of children and adolescents. The Child and Adolescent Psychiatric Assessment (CAPA) may be used to measure depression and anxiety. The impact on self-harm and suicidality (ideation and behaviour) may be measured using the Suicide Ideation Questionnaire Junior. Other measures may be used as an alternative to the stated measure. • **Impact on Quality of Life** This outcome is critical because gender dysphoria in children and adolescents may be associated with a significant reduction in health-related quality of life. Quality of Life may be measured by the KINDL questionnaire, Kidscreen 52. Other measures as reported in studies may be used as an alternative to the stated measures. *Important to decision making* • **Impact on body image** This outcome is important because some young people with gender dysphoria may desire to take steps to suppress features of their physical appearance associated with their sex assigned at birth or accentuate physical features of their experienced gender. The Body Image Scale could be used as a measure. Other measures as reported in studies may also be used as an alternative to the stated measure. • **Psychosocial Impact** Examples of psychosocial impact are: coping mechanisms which may impact on substance misuse; family relationships; peer relationships. This outcome is important because gender dysphoria in adolescents and children is associated with internalising and externalising behaviours and emotional and behavioural problems which may impact on social and occupational functioning. The child behavioural check list (CBCL) may be used to measure the impact on psychosocial functioning. Other measures as reported in studies may be used as an alternative to the stated measure. • **Engagement with health care services** This outcome is important because patient engagement with healthcare services will impact on their clinical outcomes. Engagement with health care services may be measured using the Youth Health Care measure-satisfaction, utilization, and needs (YHC-SUN) questionnaire. Loss to follow up and

	should also be ascertained as part of this outcome. Alternative measures to the YHC-SUN questionnaire may be used as reported in studies. • **Transitioning surgery - Impact on extent of and satisfaction with surgery** This outcome is important because some children and adolescents with gender dysphoria may in adulthood proceed to transitioning surgery. Stated measures of the extent of surgery and satisfaction with surgery in studies may be reported. • **De-transition** The proportion of patients who de-transition following the commencement of gender-affirming hormone treatment and the reasons why. This outcome is important to patients because there is uncertainty about the short and long term safety and adverse effects of gender-affirming hormones in children and adolescents with gender dysphoria. **B: Safety** • Short and long -term safety and adverse effects of taking gender-affirming hormones is important to assess whether treatment causes acute side effects that may lead to withdrawing the treatment or long term effects that may impact on decisions for transitioning or de-transitioning. Aspects to be reported on should include Impact of the drug use such as clinically relevant derangement in renal and liver function tests, lipids, glucose, insulin and glycosylated haemoglobin, cognitive development and functioning. The clinical and physical impact of temporary and permanent withdrawal the drug such as when patients decide to de-transition – e.g. delay in the attainment of peak bone mass, attenuation of peak bone mass, permanent physical effects. **C: Cost effectiveness** Cost effectiveness studies should be reported.
Inclusion criteria	
Study design	Systematic reviews, randomised controlled trials, controlled clinical trials, cohort studies. If no higher level quality evidence is found, case series can be considered.
Language	English only
Patients	Human studies only
Age	18 years or less
Date limits	2000-2020

Exclusion criteria	
Publication type	Conference abstracts, non-systematic reviews, narrative reviews, commentaries, letters, editorials, guidelines and pre-publication prints
Study design	Case reports, resource utilisation studies

Appendix B Search strategy

Medline, Embase, the Cochrane Library, HTA and APA PsycInfo were searched on 21 July 2020, limiting the search to papers published in English language in the last 20 years. Conference abstracts, non-systematic reviews, narrative reviews, commentaries, letters, editorials, guidelines, pre-publication prints, case reports and resource utilisation studies were excluded.

Database: Medline
Platform: Ovid
Version: Ovid MEDLINE(R) <1946 to July 17, 2020>
Search date: 21 Jul 2020
Number of results retrieved: 650
Search strategy:
Database: Ovid MEDLINE(R) <1946 to July 17, 2020>
Search Strategy:
---
1 Gender Dysphoria/ (485)
2 Gender Identity/ (18431)
3 "Sexual and Gender Disorders"/ (75)
4 Transsexualism/ (3758)
5 Transgender Persons/ (3134)
6 Health Services for Transgender Persons/ (136)
7 exp Sex Reassignment Procedures/ (835)
8 (gender* adj3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (7223)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (12665)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (102312)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (6969)
12 (male-to-female or m2f or female-to-male or f2m).tw. (114785)
13 or/1-12 (252562)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (1137237)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (852126)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (1912796)
17 Minors/ (2572)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (2360626)
19 exp pediatrics/ (58102)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (835833)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (2023650)
22 Puberty/ (13277)

23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (424041)
24 Schools/ (38087)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (7199)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (468784)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (89314)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (887443)
29 or/14-28 (5532185)
30 13 and 29 (79220)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (7)
32 30 or 31 (79220)
33 Hormones/ad, tu, th (4514)
34 exp Progesterone/ad, tu, th (10899)
35 exp Estrogens/ad, tu, th (28936)
36 exp Gonadal Steroid Hormones/ad, tu, th (34137)
37 (progesteron* or oestrogen* or estrogen*).tw. (196074)
38 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (544)
39 exp Estradiol/ad, tu, th (10823)
40 exp Testosterone/ad, tu, th (8318)
41 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (74936)
42 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (90464)
43 or/33-42 (304239)
44 32 and 43 (3183)
45 limit 44 to yr="2000 -Current" (2019)
46 animals/ not humans/ (4685420)
47 45 not 46 (1194)
48 limit 47 to english language (1155)
49 (MEDLINE or pubmed).tw. (163678)
50 systematic review.tw. (121198)
51 systematic review.pt. (130231)
52 meta-analysis.pt. (117148)
53 intervention$.ti. (123904)
54 or/49-53 (380217)
55 randomized controlled trial.pt. (509468)
56 randomi?ed.mp. (796957)
57 placebo.mp. (194937)
58 or/55-57 (848627)
59 exp cohort studies/ or exp epidemiologic studies/ or exp clinical trial/ or exp evaluation studies as topic/ or exp statistics as topic/ (5562241)
60 ((control and (group* or study)) or (time and factors)).mp. (3274107)
61 (program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. (4624419)
62 or/59-61 (9030680)
63 Observational Studies as Topic/ (5177)
64 Observational Study/ (81866)
65 Epidemiologic Studies/ (8358)

66 exp Case-Control Studies/ (1090891)
67 exp Cohort Studies/ (2011414)
68 Cross-Sectional Studies/ (332273)
69 Controlled Before-After Studies/ (526)
70 Historically Controlled Study/ (185)
71 Interrupted Time Series Analysis/ (913)
72 Comparative Study.pt. (1866044)
73 case control$.tw. (112152)
74 case series.tw. (59119)
75 (cohort adj (study or studies)).tw. (170281)
76 cohort analy$.tw. (6758)
77 (follow up adj (study or studies)).tw. (45131)
78 (observational adj (study or studies)).tw. (86247)
79 longitudinal.tw. (204239)
80 prospective.tw. (495367)
81 retrospective.tw. (442876)
82 cross sectional.tw. (284856)
83 or/63-82 (4368140)
84 54 or 58 or 62 or 83 (9402123)
85 48 and 84 (683)
86 limit 85 to (letter or historical article or comment or editorial or news or case reports) (33)
87 85 not 86 (650)

Database: Medline in-process
Platform: Ovid
Version: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <1946 to July 17, 2020>
Search date: 21 July 2020
Number of results retrieved: 122
Search strategy:
Database: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <1946 to July 17, 2020>
Search Strategy:
---

1 Gender Dysphoria/ (0)
2 Gender Identity/ (0)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (0)
5 Transgender Persons/ (0)
6 Health Services for Transgender Persons/ (0)
7 exp Sex Reassignment Procedures/ (0)
8 (gender* adj3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (1473)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (2315)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (20821)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (963)
12 (male-to-female or m2f or female-to-male or f2m).tw. (15453)
13 or/1-12 (39735)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (0)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (80295)

16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (0)
17 Minors/ (0)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (320315)
19 exp pediatrics/ (0)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (119124)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (0)
22 Puberty/ (0)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (59969)
24 Schools/ (0)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (0)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (68979)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (10287)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (112220)
29 or/14-28 (523053)
30 13 and 29 (9143)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (3)
32 30 or 31 (9144)
33 Hormones/ad, tu, th (0)
34 exp Progesterone/ad, tu, th (0)
35 exp Estrogens/ad, tu, th (0)
36 exp Gonadal Steroid Hormones/ad, tu, th (0)
37 (progesteron* or oestrogen* or estrogen*).tw. (13291)
38 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (241)
39 exp Estradiol/ad, tu, th (0)
40 exp Testosterone/ad, tu, th (0)
41 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (5458)
42 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (4772)
43 or/33-42 (19706)
44 32 and 43 (316)
45 limit 44 to yr="2000 -Current" (303)
46 animals/ not humans/ (1)
47 45 not 46 (303)
48 limit 47 to english language (303)
49 (MEDLINE or pubmed).tw. (36030)
50 systematic review.tw. (29830)
51 systematic review.pt. (1007)
52 meta-analysis.pt. (49)
53 intervention$.ti. (21354)
54 or/49-53 (68976)
55 randomized controlled trial.pt. (277)
56 randomi?ed.mp. (74978)
57 placebo.mp. (18290)
58 or/55-57 (81427)
59 exp cohort studies/ or exp epidemiologic studies/ or exp clinical trial/ or exp evaluation studies as topic/ or exp statistics as topic/ (455)

60 ((control and (group* or study)) or (time and factors)).mp. (214372)
61 (program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. (339764)
62 or/59-61 (507046)
63 Observational Studies as Topic/ (0)
64 Observational Study/ (91)
65 Epidemiologic Studies/ (0)
66 exp Case-Control Studies/ (1)
67 exp Cohort Studies/ (1)
68 Cross-Sectional Studies/ (0)
69 Controlled Before-After Studies/ (0)
70 Historically Controlled Study/ (0)
71 Interrupted Time Series Analysis/ (0)
72 Comparative Study.pt. (46)
73 case control$.tw. (14451)
74 case series.tw. (13070)
75 (cohort adj (study or studies)).tw. (29119)
76 cohort analy$.tw. (1039)
77 (follow up adj (study or studies)).tw. (3540)
78 (observational adj (study or studies)).tw. (17421)
79 longitudinal.tw. (34485)
80 prospective.tw. (63689)
81 retrospective.tw. (73761)
82 cross sectional.tw. (60195)
83 or/63-82 (250805)
84 54 or 58 or 62 or 83 (687622)
85 48 and 84 (126)
86 limit 85 to (letter or historical article or comment or editorial or news or case reports) (4)
87 85 not 86 (122)

Database: Medline epubs ahead of print
Platform: Ovid
Version: Ovid MEDLINE(R) Epub Ahead of Print <July 17, 2020>
Search date: 21 July 2020
Number of results retrieved: 32
Search strategy:
Database: Ovid MEDLINE(R) Epub Ahead of Print <July 17, 2020>
Search Strategy:
---

1 Gender Dysphoria/ (0)
2 Gender Identity/ (0)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (0)
5 Transgender Persons/ (0)
6 Health Services for Transgender Persons/ (0)
7 exp Sex Reassignment Procedures/ (0)
8 (gender* adj3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (430)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (637)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (1499)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (179)
12 (male-to-female or m2f or female-to-male or f2m).tw. (2460)

13 or/1-12 (4883)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (0)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (15416)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (0)
17 Minors/ (0)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (53285)
19 exp pediatrics/ (0)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (22649)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (0)
22 Puberty/ (0)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (13005)
24 Schools/ (0)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (0)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (12420)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (1407)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (20083)
29 or/14-28 (87968)
30 13 and 29 (1618)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (1)
32 30 or 31 (1618)
33 Hormones/ad, tu, th (0)
34 exp Progesterone/ad, tu, th (0)
35 exp Estrogens/ad, tu, th (0)
36 exp Gonadal Steroid Hormones/ad, tu, th (0)
37 (progesteron* or oestrogen* or estrogen*).tw. (1876)
38 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (63)
39 exp Estradiol/ad, tu, th (0)
40 exp Testosterone/ad, tu, th (0)
41 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (846)
42 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (665)
43 or/33-42 (2850)
44 32 and 43 (64)
45 limit 44 to yr="2000 -Current" (61)
46 animals/ not humans/ (0)
47 45 not 46 (61)
48 limit 47 to english language (61)
49 (MEDLINE or pubmed).tw. (7948)
50 systematic review.tw. (7508)
51 systematic review.pt. (28)
52 meta-analysis.pt. (37)
53 intervention$.ti. (4267)
54 or/49-53 (15048)
55 randomized controlled trial.pt. (1)

56 randomi?ed.mp. (14113)
57 placebo.mp. (3097)
58 or/55-57 (15128)
59 exp cohort studies/ or exp epidemiologic studies/ or exp clinical trial/ or exp evaluation studies as topic/ or exp statistics as topic/ (34)
60 ((control and (group* or study)) or (time and factors)).mp. (31615)
61 (program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. (65735)
62 or/59-61 (88222)
63 Observational Studies as Topic/ (0)
64 Observational Study/ (4)
65 Epidemiologic Studies/ (0)
66 exp Case-Control Studies/ (0)
67 exp Cohort Studies/ (0)
68 Cross-Sectional Studies/ (0)
69 Controlled Before-After Studies/ (0)
70 Historically Controlled Study/ (0)
71 Interrupted Time Series Analysis/ (0)
72 Comparative Study.pt. (0)
73 case control$.tw. (2577)
74 case series.tw. (2480)
75 (cohort adj (study or studies)).tw. (7959)
76 cohort analy$.tw. (287)
77 (follow up adj (study or studies)).tw. (632)
78 (observational adj (study or studies)).tw. (3763)
79 longitudinal.tw. (7079)
80 prospective.tw. (12148)
81 retrospective.tw. (16600)
82 cross sectional.tw. (9459)
83 or/63-82 (48534)
84 54 or 58 or 62 or 83 (119752)
85 48 and 84 (32)
86 limit 85 to (letter or historical article or comment or editorial or news or case reports) (0)
87 85 not 86 (32)

Database: Medline daily update
Platform: Ovid
Version: Ovid MEDLINE(R) Daily Update <July 21, 2020>
Search date: 22 July 2020
Number of results retrieved: 3
Search strategy

Database: Ovid MEDLINE(R) Daily Update <July 21, 2020>
Search Strategy:
---
1 Gender Dysphoria/ (4)
2 Gender Identity/ (38)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (2)
5 Transgender Persons/ (26)
6 Health Services for Transgender Persons/ (1)
7 exp Sex Reassignment Procedures/ (3)
8 (gender* adj3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (22)

9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (39)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (87)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (15)
12 (male-to-female or m2f or female-to-male or f2m).tw. (181)
13 or/1-12 (358)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (932)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (981)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (1756)
17 Minors/ (3)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (3672)
19 exp pediatrics/ (75)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (1658)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (2006)
22 Puberty/ (8)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (732)
24 Schools/ (56)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (5)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (622)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (98)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (1301)
29 or/14-28 (6705)
30 13 and 29 (130)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (0)
32 30 or 31 (130)
33 Hormones/ad, tu, th (3)
34 exp Progesterone/ad, tu, th (3)
35 exp Estrogens/ad, tu, th (8)
36 exp Gonadal Steroid Hormones/ad, tu, th (22)
37 (progesteron* or oestrogen* or estrogen*).tw. (161)
38 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (3)
39 exp Estradiol/ad, tu, th (8)
40 exp Testosterone/ad, tu, th (8)
41 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (79)
42 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (61)
43 or/33-42 (261)
44 32 and 43 (7)
45 limit 44 to yr="2000 -Current" (7)
46 animals/ not humans/ (3647)
47 45 not 46 (6)
48 limit 47 to english language (6)
49 (MEDLINE or pubmed).tw. (529)
50 systematic review.tw. (512)

51 systematic review.pt. (522)
52 meta-analysis.pt. (370)
53 intervention$.ti. (247)
54 or/49-53 (1065)
55 randomized controlled trial.pt. (595)
56 randomi?ed.mp. (1203)
57 placebo.mp. (219)
58 or/55-57 (1234)
59 exp cohort studies/ or exp epidemiologic studies/ or exp clinical trial/ or exp evaluation studies as topic/ or exp statistics as topic/ (7958)
60 ((control and (group* or study)) or (time and factors)).mp. (4307)
61 (program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. (5828)
62 or/59-61 (11814)
63 Observational Studies as Topic/ (27)
64 Observational Study/ (449)
65 Epidemiologic Studies/ (7)
66 exp Case-Control Studies/ (2173)
67 exp Cohort Studies/ (3287)
68 Cross-Sectional Studies/ (837)
69 Controlled Before-After Studies/ (1)
70 Historically Controlled Study/ (0)
71 Interrupted Time Series Analysis/ (6)
72 Comparative Study.pt. (768)
73 case control$.tw. (182)
74 case series.tw. (139)
75 (cohort adj (study or studies)).tw. (561)
76 cohort analy$.tw. (22)
77 (follow up adj (study or studies)).tw. (40)
78 (observational adj (study or studies)).tw. (253)
79 longitudinal.tw. (429)
80 prospective.tw. (778)
81 retrospective.tw. (1032)
82 cross sectional.tw. (739)
83 or/63-82 (5471)
84 54 or 58 or 62 or 83 (12581)
85 48 and 84 (3)
86 limit 85 to (letter or historical article or comment or editorial or news or case reports) (0)
87 85 not 86 (3)

Database: Embase
Platform: Ovid
Version: Embase <1974 to 2020 July 22>
Search date: 23 July 2020
Number of results retrieved: 1207
Search strategy:

Database: Embase <1974 to 2020 July 22>
Search Strategy:
---
1 exp Gender Dysphoria/ (5399)
2 Gender Identity/ (16820)
3 "Sexual and Gender Disorders"/ (24689)
4 Transsexualism/ (3869)
5 exp Transgender/ (6597)

6 Health Services for Transgender Persons/ (158848)
7 exp Sex Reassignment Procedures/ (1108)
8 (gender* adj3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (12470)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (22509)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (154446)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (10327)
12 (male-to-female or m2f or female-to-male or f2m).tw. (200166)
13 or/1-12 (581748)
14 exp juvenile/ or Child Behavior/ or Child Welfare/ or Child Health/ or infant welfare/ or "minor (person)"/ or elementary student/ or adolescent health/ or middle school student/ or high school student/ (3440943)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (1186161)
16 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (3586795)
17 exp pediatrics/ (106214)
18 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (1491597)
19 exp adolescence/ or exp adolescent behavior/ or adolescent health/ or high school student/ or middle school student/ (105108)
20 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (641660)
21 school/ or high school/ or kindergarten/ or middle school/ or primary school/ or nursery school/ or day care/ (103791)
22 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (687437)
23 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (138908)
24 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (1562903)
25 or/14-24 (7130881)
26 13 and 25 (181778)
27 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (17)
28 26 or 27 (181778)
29 hormone/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (5160)
30 exp progesterone derivative/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (23479)
31 exp estrogen/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (57641)
32 steroid hormone/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (372)
33 sex hormone/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (1984)
34 hormonal therapy/ (42222)
35 (progesteron* or oestrogen* or estrogen*).tw. (254142)
36 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (1224)
37 exp estradiol derivative/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (30740)

38 exp testosterone derivative/bd, ad, an, cr, do, it, dt, to, ei, ih, ia, ar, cv, dl, im, na, ip, ut, va, iv, ve, vi, po, pa, pr, sc, li, th, tp, td (15868)
39 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (99596)
40 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (114290)
41 or/29-40 (438737)
42 28 and 41 (6053)
43 limit 42 to yr="2000 -Current" (4741)
44 nonhuman/ not human/ (4649157)
45 43 not 44 (3636)
46 limit 45 to english language (3513)
47 (MEDLINE or pubmed).tw. (261145)
48 exp systematic review/ or systematic review.tw. (302985)
49 meta-analysis/ (191173)
50 intervention$.ti. (200041)
51 or/47-50 (660206)
52 random:.tw. (1552336)
53 placebo:.mp. (455979)
54 double-blind:.tw. (210671)
55 or/52-54 (1807280)
56 cohort analysis/ (596360)
57 exp epidemiology/ (3434332)
58 exp clinical trial/ (1504711)
59 evaluation study/ (45870)
60 statistics/ (301181)
61 ((control and (group* or study)) or (time and factors)).mp. (3324555)
62 (program or survey* or ci or cohort or comparative stud* or evaluation studies or follow-up*).mp. (6067112)
63 or/56-62 (11048972)
64 Clinical study/ (155444)
65 Case control study/ (157943)
66 Family study/ (26047)
67 Longitudinal study/ (141660)
68 Retrospective study/ (937696)
69 comparative study/ (859061)
70 Prospective study/ (613138)
71 Randomized controlled trials/ (182542)
72 70 not 71 (606604)
73 Cohort analysis/ (596360)
74 cohort analy$.tw. (13020)
75 (Cohort adj (study or studies)).tw. (302159)
76 (Case control$ adj (study or studies)).tw. (137432)
77 (follow up adj (study or studies)).tw. (63423)
78 (observational adj (study or studies)).tw. (168428)
79 (epidemiologic$ adj (study or studies)).tw. (106448)
80 (cross sectional adj (study or studies)).tw. (220073)
81 case series.tw. (104089)
82 prospective.tw. (861922)
83 retrospective.tw. (886445)
84 or/64-69,72-83 (4047788)
85 51 or 55 or 63 or 84 (12494560)
86 46 and 85 (2151)
87 86 not (letter or editorial).pt. (2137)

88 87 not (conference abstract or conference paper or conference proceeding or "conference review").pt. (1207)

Database: APA PsycInfo

Platform: Ovid
Version: APA PsycInfo <1806 to July Week 2 2020>
Search date: 22 July 2020
Number of results retrieved: 581
Search strategy:

Database: APA PsycInfo <1806 to July Week 2 2020>
Search Strategy:
---
1 Gender Dysphoria/ (936)
2 Gender Identity/ (8648)
3 Transsexualism/ (2825)
4 Transgender/ (5257)
5 exp Gender Reassignment/ (568)
6 (gender* adj3 (dysphori* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (15276)
7 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (13028)
8 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (7679)
9 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (5796)
10 (male-to-female or m2f or female-to-male or f2m).tw. (63688)
11 or/1-10 (99498)
12 exp Infant Development/ (21841)
13 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (150219)
14 Child Characteristics/ or exp Child Behavior/ or Child Psychology/ or exp Child Welfare/ or Child Psychiatry/ (23423)
15 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (984230)
16 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (78962)
17 Adolescent Psychiatry/ or Adolescent Behavior/ or Adolescent Development/ or Adolescent Psychology/ or Adolescent Characteristics/ or Adolescent Health/ (62142)
18 Puberty/ (2753)
19 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (347604)
20 Schools/ (29181)
21 Child Day Care/ or Nursery Schools/ (2836)
22 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (772814)
23 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (21475)
24 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (285697)
25 or/12-24 (1765408)
26 11 and 25 (49560)
27 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (14)

28 26 or 27 (49561)
29 hormones/ (8408)
30 sex hormones/ (1777)
31 exp progestational hormones/ (2409)
32 estrogens/ (3889)
33 steroids/ (3797)
34 (progesteron* or oestrogen* or estrogen*).tw. (11188)
35 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. (457)
36 estradiol/ (3120)
37 testosterone/ (5606)
38 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. (9625)
39 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. (6741)
40 or/29-39 (30344)
41 28 and 40 (1005)
42 limit 41 to yr="2000 -Current" (749)
43 limit 42 to english language (692)
44 limit 43 to ("0200 book" or "0240 authored book" or "0280 edited book" or "0300 encyclopedia" or "0400 dissertation abstract") (111)
45 43 not 44 (581)

Database: Cochrane Library – incorporating Cochrane Database of Systematic Reviews (CDSR); CENTRAL

Platform: Wiley
Version:
CDSR –Issue 7 of 12, July 2020
CENTRAL – Issue 7 of 12, July 2020
Search date: 22 July 2020
Number of results retrieved: CDSR 0 ; CENTRAL 67.

ID Search Hits
#1 MeSH descriptor: [Gender Dysphoria] this term only 3
#2 MeSH descriptor: [Gender Identity] this term only 227
#3 MeSH descriptor: [Sexual and Gender Disorders] this term only 2
#4 MeSH descriptor: [Transsexualism] this term only 27
#5 MeSH descriptor: [Transgender Persons] this term only 36
#6 MeSH descriptor: [Health Services for Transgender Persons] this term only 0
#7 MeSH descriptor: [Sex Reassignment Procedures] explode all trees 4
#8 (gender* near/3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*)):ti,ab,kw 702
#9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*):ti,ab,kw 959
#10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*):ti,ab,kw 3969
#11 ((sex or gender*) near/3 (reassign* or chang* or transform* or transition*)):ti,ab,kw 524
#12 (male-to-female or m2f or female-to-male or f2m):ti,ab,kw 516
#13 #1 or #2 or #3 or #4 or #5 or #6 or #7 or #8 or #9 or #10 or #11 or #12 6413
#14 MeSH descriptor: [Infant] explode all trees 28440
#15 MeSH descriptor: [Infant Health] this term only 49
#16 MeSH descriptor: [Infant Welfare] this term only 82

#17 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*):ti,ab,kw,so 89530
#18 MeSH descriptor: [Child] explode all trees 44089
#19 MeSH descriptor: [Child Behavior] explode all trees 2061
#20 MeSH descriptor: [Child Health] this term only 98
#21 MeSH descriptor: [Child Welfare] this term only 325
#22 MeSH descriptor: [Minors] this term only 8
#23 (child* or minor or minors or boy* or girl* or kid or kids or young*):ti,ab,kw,so 265417
#24 MeSH descriptor: [Pediatrics] explode all trees 661
#25 (pediatric* or paediatric* or peadiatric*):ti,ab,kw,so 57725
#26 MeSH descriptor: [Adolescent] this term only 102154
#27 MeSH descriptor: [Adolescent Behavior] this term only 1358
#28 MeSH descriptor: [Adolescent Health] this term only 29
#29 MeSH descriptor: [Puberty] this term only 295
#30 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*):ti,ab,kw,so 140927
#31 MeSH descriptor: [Schools] this term only 1914
#32 MeSH descriptor: [Child Day Care Centers] this term only 231
#33 MeSH descriptor: [Nurseries, Infant] explode all trees 17
#34 MeSH descriptor: [Schools, Nursery] this term only 37
#35 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*):ti,ab,kw,so 97810
#36 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") near/2 (year or years or age or ages or aged)):ti,ab 6710
#37 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") near/2 (year or years or age or ages or aged)):ti,ab 196881
#38 #14 or #15 or #16 or #17 or #18 or #19 or #20 or #21 or #22 or #23 or #24 or #25 or #26 or #27 or #28 or #29 or #30 or #31 or #32 or #33 or #34 or #35 or #36 or #37 516067
#39 #13 and #38 2488
#40 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*):ti,ab,kw 0
#41 #39 or #40 2488
#42 MeSH descriptor: [Hormones] this term only 2241
#43 MeSH descriptor: [Progesterone] explode all trees 3135
#44 MeSH descriptor: [Estrogens] explode all trees 1841
#45 MeSH descriptor: [Gonadal Steroid Hormones] explode all trees 10747
#46 (progesteron* or oestrogen* or estrogen*):ti,ab,kw 18387
#47 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)):ti,ab,kw 24
#48 MeSH descriptor: [Estradiol] explode all trees 4434
#49 MeSH descriptor: [Testosterone] explode all trees 2945
#50 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan):ti,ab,kw 7386
#51 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle):ti,ab,kw 11410
#52 #42 or #43 or #44 or #45 or #46 or #47 or #48 or #49 or #50 or #51 31870
#53 #41 and #52 121
#54 "conference":pt or (clinicaltrials or trialsearch):so 492465
#55 #53 not #54 72

Database: HTA

Platform: Wiley
Version: up to 2018
Search date: 22nd July 2020
Number of results retrieved: 4
Search strategy:

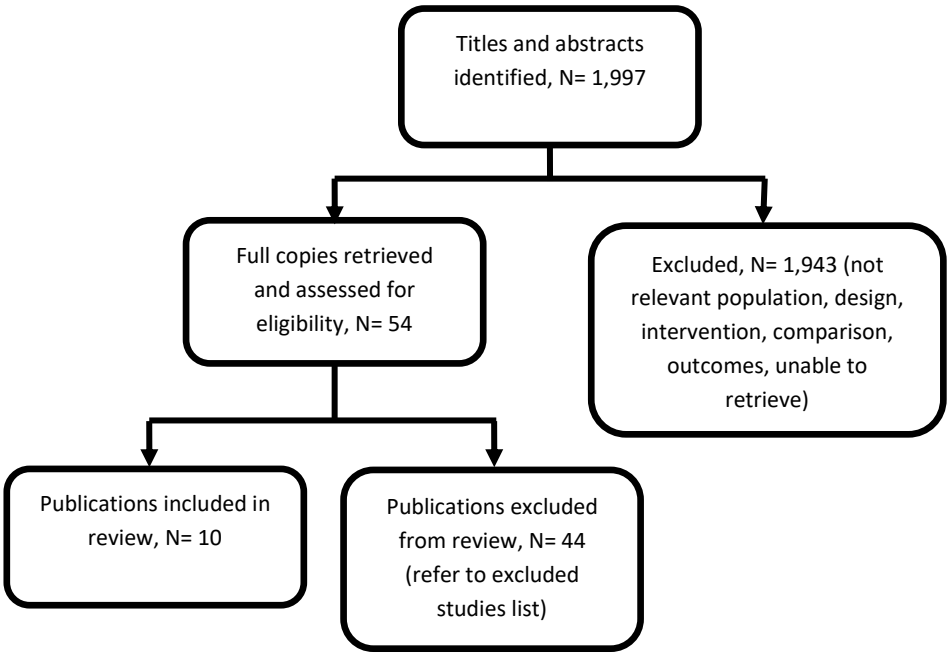
#1 MeSH DESCRIPTOR Gender Dysphoria 0
#2 MeSH DESCRIPTOR Gender Identity 12
#3 MeSH DESCRIPTOR Sexual and Gender Disorders 2
#4 MeSH DESCRIPTOR Transsexualism 12
#5 MeSH DESCRIPTOR Transgender Persons 3
#6 MeSH DESCRIPTOR Health Services for Transgender Persons 0
#7 MeSH DESCRIPTOR Sex Reassignment Procedures EXPLODE ALL TREES 1
#8 ((gender* near3 (dysphori* or incongru* or identi* or disorder* or confus* or minorit* or queer*))) 28
#9 ((transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*)) 76
#10 ((trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*)) 83
#11 (((sex or gender*) near3 (reassign* or chang* or transform* or transition*))) 24
#12 ((male-to-female or m2f or female-to-male or f2m)) 86
#13 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 261
#14 MeSH DESCRIPTOR Infant EXPLODE ALL TREES 2964
#15 MeSH DESCRIPTOR Infant Health 0
#16 MeSH DESCRIPTOR Infant Welfare 22
#17 ((prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*)) 5510
#18 MeSH DESCRIPTOR Child EXPLODE ALL TREES 4935
#19 MeSH DESCRIPTOR Child Behavior EXPLODE ALL TREES 64
#20 MeSH DESCRIPTOR Child Health 2
#21 MeSH DESCRIPTOR Child Welfare 80
#22 MeSH DESCRIPTOR Minors 2
#23 ((child* or minor or minors or boy* or girl* or kid or kids or young*)) 13575
#24 MeSH DESCRIPTOR Pediatrics EXPLODE ALL TREES 119
#25 ((pediatric* or paediatric* or peadiatric*)) 2842
#26 MeSH DESCRIPTOR Adolescent 4594
#27 MeSH DESCRIPTOR Adolescent Behavior 94
#28 MeSH DESCRIPTOR Adolescent Health 0
#29 MeSH DESCRIPTOR Puberty 3
#30 ((adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*)) 5621
#31 MeSH DESCRIPTOR Schools 168
#32 MeSH DESCRIPTOR Child Day Care Centers 12
#33 MeSH DESCRIPTOR Schools, Nursery 3
#34 ((pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*)) 4454
#35 ((("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") near2 (year or years or age or ages or aged))) 380
#36 ((("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") near2 (year or years or age or ages or aged))) 7996

#37 #14 OR #15 OR #16 OR #17 OR #18 OR #19 OR #20 OR #21 OR #22 OR #23 OR #24 OR #25 OR #26 OR #27 OR #28 OR #29 OR #30 OR #31 OR #32 OR #33 OR #34 OR #35 OR #36 22640
#38 #13 AND #37 116
#39 (#13 AND #37) IN HTA 4

Appendix C Evidence selection

The literature searches identified 1,997 references. These were screened using their titles and abstracts and 54 references were obtained and assessed for relevance. Of these, 10 references are included in the evidence review. The remaining 44 references were excluded and are listed in appendix D.

Figure 1 – Study selection flow diagram



References submitted with Preliminary Policy Proposal

There is no preliminary policy proposal for this policy.

Appendix D Excluded studies table

Study reference	Reason for exclusion
Aranda G, Mora M, Hanzu FA et al. (2019) Effects of sex steroids on cardiovascular risk profile in transgender men under gender affirming hormone therapy. Endocrinologia, diabetes y nutricion 66(6): 385–392	Excluded on population – adult study, participants not 18 years or less (mean age 27.1 years).
Arnold, Justin D, Sarkodie, Eleanor P, Coleman, Megan E et al. (2016) Incidence of Venous Thromboembolism in Transgender Women	Excluded on population – adult study, participants not 18 years or less (mean age 33.2 years).

Study reference	Reason for exclusion
Receiving Oral Estradiol. The journal of sexual medicine 13(11): 1773–1777	
Asscheman, Henk, Giltay, Erik J, Megens, Jos A J et al. (2011) A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones. European journal of endocrinology 164(4): 635–42	Excluded on population – although some participants started gender-affirming hormones when young, the study does not report the proportion who started treatment when 18 years or less. Mean ages at start of treatment were 31.4 years (transfemales) and 26.1 years (transmales), suggesting the majority of participants were older than 18 years at the start of treatment. Outcomes not reported separately for people aged 18 years or less.
Author not, found (2014) Hormone therapy for the treatment of gender dysphoria. Lansdale, PA: HAYES, Inc	Full text paper not available.
Baba, T., Endo, T., Honnma, H. et al. (2007) Association between polycystic ovary syndrome and female-to-male transsexuality. Human Reproduction 22(4): 1011–1016	Excluded on population – although study included some younger people (age range 17 to 47), most participants were adults (mean age around 25 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.
Becerra-Fernandez A, Perez-Lopez G, Roman MM et al. (2014) Prevalence of hyperandrogenism and polycystic ovary syndrome in female to male transsexuals. Endocrinologia y Nutricion: Organo de la Sociedad Espanola de Endocrinologia y Nutricion 61(7): 351–8	Excluded on population – although study included some younger people (age range 18 to 45), most participants were adults (mean age around 25 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.
Becker I, Auer M, Barkmann C et al. (2018) A Cross-Sectional Multicenter Study of Multidimensional Body Image in Adolescents and Adults with Gender Dysphoria Before and After Transition-Related Medical Interventions. Archives of Sexual Behavior 47(8): 2335–2347	Excluded on population – study included people aged 14 to 21 years. Outcomes not reported separately for people aged 18 years or less. Better evidence available – only 11 participants received gender-affirming hormones. The majority of the study cohort were either pre-treatment, received puberty suppression alone, or received hormones and underwent surgery.
Chew D, Anderson J, Williams K et al. (2018) Hormonal Treatment in Young People With Gender Dysphoria: A Systematic Review. Pediatrics 141(4): e20173742	Excluded on better available evidence - systematic review did not meta-analyse results from. Individual studies from this systematic review are either

Study reference	Reason for exclusion
	included, or excluded because they did not meet the PICO criteria.
Connolly MD, Zervos MJ, Barone CJ 2nd et al. (2016) The Mental Health of Transgender Youth: Advances in Understanding. The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine 59(5): 489–495	Excluded on intervention - review did not investigate gender-affirming hormones
de Vries ALC, McGuire JK, Steensma TD et al. (2014) Young adult psychological outcome after puberty suppression and gender reassignment. Pediatrics 134(4): 696–704	Exclude on intervention – all participants had surgery after gender-affirming hormones. Unable to determine whether changes were due to hormones or surgery. Complete data only available for 40 patients. Details of gender-affirming hormones are poorly reported. Outcomes reported in other study (with a population that more closely matches PICO)
Elamin MB, Garcia MZ, Murad MH et al. (2010) Effect of sex steroid use on cardiovascular risk in transsexual individuals: a systematic review and meta-analyses. Clinical Endocrinology 72(1): 1–10	Exclude on population – all included studies conducted in adult population. Unclear whether hormones were started when participants were aged 18 years or less. Outcomes not reported by age at treatment initiation.
Fernandez JD and Tannock LR (2016) Metabolic effects of hormone therapy in transgender patients. Endocrine Practice: Official Journal of the American College of Endocrinology and the American Association of Clinical Endocrinologists 22(4): 383–8	Excluded on population – adult study, participants not 18 years or less (mean ages 31 and 27 years).
Fighera TM, Ziegelmann PK, Da Silva TR et al. (2019) Bone mass effects of cross-sex hormone therapy in transgender people: Updated systematic review and meta-analysis. Journal of the Endocrine Society 3(5): 943–964	Excluded on population – all included studies conducted in adult population. Unclear whether hormones were started when participants were aged 18 years or less. Outcomes not reported by age at treatment initiation.
Getahun D, Nash R, Flanders WD et al. (2018) Cross-sex Hormones and Acute Cardiovascular Events in Transgender Persons: A Cohort Study. Annals of Internal Medicine 169(4): 205–213	Excluded on population – adult study, participants not 18 years or less.
Gomez-Gil E, Zubiaurre-Elorza L, de Antonio IE et al. (2014) Determinants of quality of life in Spanish transsexuals attending a gender unit before genital sex reassignment surgery. Quality of Life Research: an International Journal of Quality of Life Aspects of Treatment, Care and Rehabilitation 23(2): 669–76	Excluded on population – although study included some younger people (age range 16 to 67), most participants were adults (mean age 31.2 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.
Gomez-Gil E, Zubiaurre-Elorza L, Esteva I et al. (2012) Hormone-treated transsexuals report less	Excluded on population – adult study, participants not 18 years or less (mean age 24.6 years).

Study reference	Reason for exclusion
social distress, anxiety and depression. Psychoneuroendocrinology 37(5): 662–70	
Gooren LJ, van Trotsenburg MAA, Giltay EJ et al. (2013) Breast cancer development in transsexual subjects receiving cross-sex hormone treatment. The Journal of Sexual Medicine 10(12): 3129–34	Excluded on population – study reports on cancer rates in people aged 18-80 years. The 3 cases of cancer all started gender-affirming hormone treatment >18 years.
Grimstad FW, Boskey E, Grey M (2020) New-Onset Abdominopelvic Pain After Initiation of Testosterone Therapy Among TransMasculine Persons: A Community-Based Exploratory Survey. LGBT health 7(5): Published Online:13 Jul 2020https://doi.org/10.1089/lgbt.2019.0258	Excluded on population – adult study, participants not 18 years or less.
Hannema SE, Schagen SEE, Cohen-Kettenis PT et al. (2017) Efficacy and Safety of Pubertal Induction Using 17beta-Estradiol in Transgirls. The Journal of Clinical Endocrinology and Metabolism 102(7): 2356–2363	Excluded on better evidence available – small study (n=28) with high drop-out rate (n=16 at final follow-up). Same outcomes reported in larger studies.
Jarin J, Pine-Twaddell E, Trotman G et al. (2017) Cross-Sex Hormones and Metabolic Parameters in Adolescents With Gender Dysphoria. Pediatrics 139(5)	Excluded on population and better evidence available. Although the study included some younger people (age range 13 to 25; mean age 16 and 18), the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less. Outcomes were limited to physiological results (including haemoglobin, lipids and BMI). Follow-up only 6 months, other included studies report same outcomes with longer follow-up (12 to 31 months).
Keo-Meier CL, Herman LI, Reisner SL et al. (2015) Testosterone treatment and MMPI-2 improvement in transgender men: a prospective controlled study. Journal of consulting and clinical psychology 83(1): 143–56	Excluded on population – although study included some younger people (age range 18 to 54), most participants were adults (mean age 26.6 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.
Klaver M, de Mutsert R, Wiepjes CM et al. (2018) Early Hormonal Treatment Affects Body Composition and Body Shape in Young Transgender Adolescents. The Journal of Sexual Medicine 15(2): 251–260	Excluded on outcomes – reported outcomes not included in PICO document. The risk of obesity with gender-affirmed hormones was reported in an included study.
McFarlane T, Zajac JD, Cheung AS (2018) Gender-affirming hormone therapy and the risk of sex hormone-dependent tumours in transgender individuals-A systematic review. Clinical Endocrinology 89(6): 700-711	Exclude on population – all included studies conducted in adult population.

Study reference	Reason for exclusion
Meriggiola MC, Armillotta F, Costantino A et al. (2008) Effects of testosterone undecanoate administered alone or in combination with letrozole or dutasteride in female to male transsexuals. The Journal of Sexual Medicine 5(10): 2442–53	Excluded on population – adult study, participants not 18 years or less.
Nota NM, Wiepjes CM, de Blok, CJM et al. (2018) The occurrence of benign brain tumours in transgender individuals during cross-sex hormone treatment. Brain: A Journal of Neurology 141(7): 2047–2054	Excluded on population – adult study, participants not 18 years or less.
Oda H and Kinoshita T (2017) Efficacy of hormonal and mental treatments with MMPI in FtM individuals: Cross-sectional and longitudinal studies. BMC Psychiatry 17(1): 256	Excluded on population – although study included some younger people (age range 15 to 43), most participants were adults (mean age around 25.6 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.
Olson-Kennedy J, Okonta V, Clark LF et al. (2018) Physiologic Response to Gender-Affirming Hormones Among Transgender Youth. The Journal of Adolescent Health: Official Publication of the Society for Adolescent Medicine 62(4): 397–401	Excluded on population – although study included some younger people (age range 12 to 23; mean age 18 years). Outcomes not reported separately for people aged 18 years or less. Outcomes limited to physiological results (including haemoglobin, lipids, liver enzymes and BMI). Same outcomes reported in included studies that had a less indirect population and a longer follow-up.
Ott J, Kaufmann U, Bentz K et al. (2010) Incidence of thrombophilia and venous thrombosis in transsexuals under cross-sex hormone therapy. Fertility and sterility 93(4): 1267–72	Excluded on population – adult study, participants not 18 years or less.
Pakpoor J, Wotton CJ, Schmierer K et al. (2016) Gender identity disorders and multiple sclerosis risk: A national record-linkage study. Multiple sclerosis. Multiple Sclerosis Journal. 22(13): 1759–1762	Excluded on population – although study included some younger people, outcomes not reported separately for people aged 18 years or less. Also exclude for intervention – unclear if people received gender-affirming hormones.
Pyra M, Casimiro I, Rusie L et al. (2020) An Observational Study of Hypertension and Thromboembolism among Transgender Patients Using Gender-Affirming Hormone Therapy. Transgender Health 5(1): 1–9	Excluded on population – adult study (age range 20-70). Age at which gender-affirming hormones started not reported.
Quiros C, Patrascioiu I, Mora M et al. (2015) Effect of cross-sex hormone treatment on cardiovascular risk factors in transsexual individuals. Experience in a specialized unit in Catalonia. Endocrinologia y nutricion : organo de la Sociedad Espanola de Endocrinologia y Nutricion 62(5): 210–6	Excluded on population – adult study, participants not 18 years or less.

Study reference	Reason for exclusion
Rowniak S, Bolt L, Sharifi C (2019) Effect of cross-sex hormones on the quality of life, depression and anxiety of transgender individuals: A quantitative systematic review. JBI Database of Systematic Reviews and Implementation Reports 17(9): 1826–1854	Exclude on population – all included studies conducted in adult population.
Sequeira GM, Kidd K, El Nokali NE et al. (2019) Early Effects of Testosterone Initiation on Body Mass Index in Transmasculine Adolescents. Journal of Adolescent Health 65(6): 818–820	Exclude on outcome - study only reports BMI z-score over 12 month testosterone treatment. BMI not listed as an outcome of interest in the PICO document. Other included studies have investigated the impact of gender-affirming hormone treatment on CV risk profile, including longer term obesity rates, with a longer follow-up and more participants.
Shim JY, Laufer MR, Grimstad FW (2020) Dysmenorrhea and Endometriosis in Transgender Adolescents. Journal of Pediatric and Adolescent Gynecology. Available online 11 June 2020. https://doi.org/10.1016/j.jpag.2020.06.001	Exclude on population – only 2 participants taking testosterone before diagnosis of dysmenorrhea.
Slabbekoorn D, Van Goozen SHM, Gooren, LJG et al. (2001) Effects of cross-sex hormone treatment on emotionality in transsexuals. International Journal of Transgenderism 5(3): http://www.symposion.com/ijt/ijtvo05no03_02.htm	Excluded on population – adult study (age range 21 to 28 years)
Smith YLS., Van Goozen SHM, Kuiper AJ et al. (2005) Sex reassignment: Outcomes and predictors of treatment for adolescent and adult transsexuals. Psychological Medicine 35(1): 89–99	Excluded on population – results on adults only used to assess hormone treatment.
Sutherland N, Espinel W, Grotzke M et al. (2020) Unanswered Questions: Hereditary breast and gynecological cancer risk assessment in transgender adolescents and young adults. Journal of Genetic Counseling 29(4): 625–633	Excluded on study type – narrative review of 3 case reports.
van Velzen DM, Paldino A, Klaver M et al. (2019) Cardiometabolic Effects of Testosterone in Transmen and Estrogen Plus Cyproterone Acetate in Transwomen. The Journal of Clinical Endocrinology and Metabolism 104(6): 1937–1947	Excluded on population – adult study, participants not 18 years or less.
White Hughto JM and Reisner SL (2016) A Systematic Review of the Effects of Hormone Therapy on Psychological Functioning and Quality of Life in Transgender Individuals. Transgender Health 1(1): 21–31	Exclude on population – all included studies conducted in adult population.
Wiepjes CM, de Blok CJM, Staphorsius AS et al. (2020) Fracture Risk in Trans Women and Trans Men Using Long-Term Gender-Affirming Hormonal Treatment: A Nationwide Cohort Study. Journal of Bone and Mineral Research 35(1): 64–70	Excluded on population – adult study, all participants started gender-affirming hormones after 18 years.
Wierckx K, Mueller S, Weyers S et al. (2012) Long-term evaluation of cross-sex hormone treatment in	Excluded on population – adult study, participants not 18 years or less.

Study reference	Reason for exclusion
transsexual persons. The Journal of Sexual Medicine 9(10): 2641–51	
Wierckx K, Van Caenegem E, Schreiner T et al. (2014) Cross-sex hormone therapy in trans persons is safe and effective at short-time follow-up: results from the European network for the investigation of gender incongruence. The journal of sexual medicine 11(8): 1999–2011	Excluded on population – adult study, participants not 18 years or less.
Wilson R, Jenkins C, Miller H et al. (2006) The effect of oestrogen on cytokine and antioxidant levels in male to female transsexual patients. Maturitas 55(1): 14–8	Excluded on population – adult study, participants not 18 years or less.
Witcomb GL, Bouman WP, Claes L et al. (2018) Levels of depression in transgender people and its predictors: Results of a large matched control study with transgender people accessing clinical services. Journal of Affective Disorders 235: 308–315	Excluded on population – although study included some younger people (age range 15 to 79), most participants were adults (mean age around 30.4 years) and the proportion who started treatment when 18 years or less is not reported. Outcomes not reported separately for people aged 18 years or less.

Appendix E Evidence tables

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Achille, C., Taggart, T., Eaton, N.R. et al. (2020) Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: Preliminary results. International Journal of Pediatric Endocrinology 2020(1): 8 **Study location** Single centre, New York, United States **Study type** Prospective longitudinal study **Study aim** To assess the psychological wellbeing and quality of life in children and adolescents who have sought endocrine	Inclusion and exclusion not reported- it appears from the description in the publication that all people referred for gender dysphoria were invited to participate, and the vast majority agreed. Of the 95 treatment naïve people who entered the study, 50 people completed all follow-up questionnaires and were included in the analysis. No description of the 45 people without follow-up data reported. The study included 50 children, adolescents and young adults with gender dysphoria.	**Intervention** Endocrine interventions (the collective term used by authors for puberty suppression and gender-affirming hormones) were introduced as per Endocrine Society and the World Professional Association for	**Critical Outcomes** ***Impact on mental health*** Depression symptoms were assessed using the Center for Epidemiologic Studies Depression Scale (CESD-R). Statistically significant improvements in CESD-R score were observed from baseline (initial assessment; 21.4 points) to about 12 months follow-up (13.9 points; p<0.001). Regression analysis, controlling for reported medicines for mental health problems and engagement in counselling, found no statistically significant change from baseline in transfemales (p=0.27) and transmales (p=0.43). The Patient Health Questionnaire Modified for Teens (PHQ 9_Modified for Teens) was also used to assess depression symptoms. Depression scores improved from baseline (p< 0.001; absolute scores not reported numerically). Regression analysis, controlling for reported medicines for mental health problems and engagement in counselling, found no statistically significant change from baseline in transfemales (p=0.07) and transmales (p=0.67). Suicidal ideation measured using the additional questions from the PHQ 9_Modified for Teens, was presented in 10% (5/50) of	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record 4. b) no **Domain 2: Comparability** 1. c) no comparator **Domain 3: Outcome** 1. c) self-report 2. a) yes – 6 monthly assessment up to 12 months (preliminary results from an ongoing study) 3. c) Follow up rate less than 80% and no description of those lost **Overall quality is assessed as poor** Other comments: Although regression analysis results for some outcomes were controlled for use of medicines for mental health problems,

intervention to help with gender dysphoria. **Study dates** Study recruitment ran from December 2013 to December 2018; study is ongoing	17 transfemales and 33 transmales. Diagnostic criteria for gender dysphoria not reported. Mean age at baseline was 16.2 years (SD 2.2). Mean age at the start of gender-affirming hormone treatment not reported.	Transgender Health (WPATH) guidelines. Puberty suppression was: • GnRH agonist and/or anti-androgens (transfemales) • GnRH agonist or medroxyprogesterone (transmales) Average duration of GnRH analogue treatment not reported. Once eligible, gender-affirming hormones were offered, these were: • Oestradiol (transfemales) • Testosterone (transmales) Doses and route of administration not reported. After about 12-months treatment ('wave 3' in the study): • 24 people (48%) were on gender-affirming hormones alone • 12 people (24%) were on puberty suppression alone	participants at baseline and 6% (3/50) at about 12-month follow-up, no statistical analysis reported. The study also reported results by gender: In transfemales, 11.8% (2/17) had suicidal ideation at baseline compared with 5.9% (1/17) at 12-month follow-up (no statistically analysis reported) In transmales, 9.1% (3/33) had suicidal ideation at baseline compared with 6.1% (2/33) at 12-month follow-up (no statistically analysis reported) ***Impact on quality of life*** Quality of Life Enjoyment and Satisfaction Questionnaire (QLES-Q-SF) scores: there was no statistically significant change in score from baseline to about 12-months (p=0.085; absolute scores not reported numerically). Regression analysis, controlling for reported medicines for mental health problems and engagement in counselling, found not statistically significant change from baseline in transfemales (p=0.06) and transmales (p=0.08). *No other critical or important outcomes reported*	details of these is not reported. Other co-morbidities not reported. Source of funding: None

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
		• 11 people (22%) were on both gender-affirming hormones and puberty suppression • 3 people (6%) were on no endocrine intervention Results not represented separately for the sub-group of people who received gender-affirming hormones. Average duration of treatment with gender-affirming hormones not reported. **Comparison** No comparison group. Change overtime reported.		

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Allen, LR, Watson, LB, Egan, AM et al. (2019) Well-being and suicidality among transgender youth after gender-affirming hormones. Clinical Practice in Pediatric	The study included adolescents and young adults (age range 13-20 years) who received services for gender dysphoria in a clinic in the United States. Participants were required to have received gender-	39 participants received gender-affirming hormones only 8 participants received a GnRH analogue followed by gender-affirming hormones.	**Critical Outcomes** ***Impact on mental health*** The Ask Suicide-Screening Questions (ASQ) instrument was used to assess suicidality. Following an average of about 12 months treatment with gender-affirming hormones, adjusted mean ASQ score was statistically significantly lower (from 1.11 [standard error	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Psychology 7(3): 302-311 **Study location** Single centre, Kansas City, United States **Study type** Retrospective longitudinal study **Study aim** To examine suicidality and general well-being following administration of gender-affirming hormones. **Study dates** Participants first presented to the clinic between 2015 and 2018.	affirming hormones for at least 3 months, and have pre-test and final assessment data points. No exclusion criteria reported. In total 47 adolescents and young adults with gender dysphoria were included: 14 transfemales (sex assigned at birth male) and 33 transmales (sex assigned at birth female). Diagnostic criteria for gender dysphoria not reported. Mean age at pre-test (before administration of gender-affirming hormones) was 16.59 years (range 13.73 to 19.04). Mean age at the start of treatment in the sub-group who received gender-affirming hormones-only was 16.72 years. Mean age at the start of treatment with gender-affirming hormones in people who previously	Mean duration of treatment in the gender-affirming hormones only subgroup was 366 days. Mean duration of gender-affirming hormone treatment in people who had previously received a GnRH analogue was not reported. Mean duration of treatment with a GnRH analogue was not reported. Participants were assessed at the start of treatment and at least 3 months after treatment.	(SE) 0.22] at baseline to 0.27 [SE 0.12] at final assessment; p<0.001). The authors also reported change in ASQ separately for transfemales (from 1.21 [SE 0.36] at baseline to 0.24 [SE 0.19] at final assessment) and transmales (from 1.01 [SE 0.36] at baseline to 0.29 [0.13] at final assessment). There was no statistically significant difference in change from baseline between transfemales and transmales (p=0.79) ***Impact on quality of life*** Assessed using the General Well-Being Scale (GWBS) of the Pediatric Quality of Life Inventory. Following an average of about 12 months treatment with gender-affirming hormones, adjusted mean GWBS score was statistically significantly higher (from 61.7 [SE 2.43] at baseline to 70.23 [2.15] at final assessment; p<0.002). The authors also reported change in GWBS of the Pediatric Quality of Life Inventory for transfemales (from 58.44 [SE 4.09] at baseline to 69.52 [SE 3.62] at final assessment) and transmales (from 64.95 [SE 2.66] at baseline to 70.94 [2.35] at final assessment). There was no statistically significant difference in change from baseline between transfemales and transmales (p=0.32) *No other critical or important outcomes reported*	3. a) secure record 4. b) no **Domain 2: Comparability** 2. c) no comparator **Domain 3: Outcome** 1. b) record linkage 2. a) yes – mean duration of treatment was 366 days 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor** Other comments: None Source of funding: Not reported

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
	received a GnRH analogue was not reported.			

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Kaltiala, R., Heino, E., Tyolajarvi, M. et al. (2020) Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria. Nordic Journal of Psychiatry 74(3): 213-219 **Study location** Single centre, Tampere, Finland **Study type** Retrospective chart review **Study aim** To evaluate the psychosocial functioning and need for mental health treatment during the gender identity diagnostic phase and after about	The study included adolescents who were referred to the gender identity service before they 18 years old, were diagnosed with gender dysphoria, received gender-affirming hormones and completed a follow-up of approximately 12 months after starting hormones. In total 52 adolescents were included, comprising of 11 transfemales and 41 transmales. Gender dysphoria was diagnosed according to International Classification of Disease 10 (ICD-10). The authors state that the corresponding diagnosis to 'gender dysphoria' in	Intervention referred to as 'hormonal sex reassignment treatment' – details of intervention not reported, although gender-affirming hormones were prescribed to all participants. It is not clear from the study whether additional interventions were prescribed. Medical records reviewed for the 'real-life phase' – the approximately 12 months follow-up period for this population in Finland.	**Critical Outcomes** ***Impact on mental health*** Of the 52 people who received gender-affirming hormones, 50% (26/52) needed mental health treatment before or during the assessment and 46% (24/51) needed mental health treatment during the 12-month 'real life' phase (no statistically significant difference). For specific symptoms / conditions: • depression: 54% (28/52) needed treatment before or during the assessment and 15% (8/52) needed treatment during the 12-month 'real life' phase (statistically significant reduction, p<0.001) • anxiety: 48% (25/52) needed treatment before or during the assessment and 15% (8/52) needed treatment during the 12-month 'real life' phase (statistically significant reduction, p<0.001) • suicidality/self-harm: 35% (18/52) needed treatment before or during the assessment and 4% (2/52) needed treatment during the 12-month 'real life' phase (statistically significant reduction, p<0.001) • conduct problems/antisocial: 14% (7/52) needed treatment before or during the	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. a) yes – 12 month follow-up 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor**

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
a year on gender-affirming hormones. **Study dates** 2011 to 2017	the ICD-10 is 'transsexualism'. Mean age at diagnosis 18.1 years (range 15.2 to 19.9)		assessment and 6% (3/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 0.18) • psychotic symptoms/psychosis: 2% (1/52) needed treatment before or during the assessment and 4% (2/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 0.56) • substance abuse: 4% (2/52) needed treatment before or during the assessment and 2% (1/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 0.56) • autism: 12% (6/52) needed treatment before or during the assessment and 6% (3/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 0.30) • ADHD: 10% (5/52) needed treatment before or during the assessment and 2% (1/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 0.09) • eating disorder: 2% (1/52) needed treatment before or during the assessment and 2% (1/52) needed treatment during the 12-month 'real life' phase (no statistically significant difference, p= 1.0). No details of actual treatment reported. **Important Outcomes** ***Psychosocial Impact*** Study reported on measures of functioning in different domains of adolescent development,	Other comments: None Source of funding: No source of funding reported

			reported over the approximately 12-month period after starting gender-affirming hormones (referred to as the 'real-life phase' in Finland) Significantly fewer participants were living with parent(s)/ guardians during the real-life phase (40%; 21/50) compared with during gender identity assessment (73%; 38/52; p=0.001)) There was a statistically significant reduction in the number of participants with normative peer contacts, from gender identity assessment (89%; 46/52) to the real-life phase (81%; 42/52; p<0.001). There was no significant difference in the number of participants who were progressing normally in school or work during gender identity assessment (64%; 33/52) compared with the real-life phase (60%; 31/52). There was no significant difference in the number of participants who have been dating or were in steady relationships during gender identity assessment (62%; 32/50) compared with the real-life phase (58%; 30/52). There was no significant difference in the number of participants who were able to deal with matters outside of the home in an age-appropriate manner during gender identity assessment (81% (42/52) compared with the real-life phase (81%; 42/52)	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			No other critical or important outcomes reported	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding

Full citation
Khatchadourian K, Amed S, Metzger DL (2014) Clinical management of youth with gender dysphoria in Vancouver. The Journal of pediatrics 164(4): 906-11
 	
Study location
Single centre study, Vancouver, Canada
 	
Study type
Retrospective chart review
 	
Study aim
To describe the patient characteristics, clinical management, and response to treatment in a cohort of people seen in a single clinic.
 	
Study dates
1998 to 2011
	Inclusion criteria were at least Tanner stage 2 pubertal development, previous assessment by a mental health professional and a confirmed diagnosis of gender dysphoria (diagnostic criteria not specified). No exclusion criteria are specified.
	
	63 children, adolescents and young people with gender dysphoria who started gender-affirming hormones, out of 84 young people seen in the unit between 1998 and 2011.
	39 transfemales and 24 transmales.
	
	Diagnostic criteria for gender dysphoria not reported.
	
	Mean age at the start of gender-affirming hormone treatment was 17.4 years (SD 1.9).
		Intervention
		Transfemales: Oestrogen (oral micronized 17β-oestradiol)
		Transmales: Testosterone (injectable testosterone enanthate and/or cypionate)
		
		19 participants (30%) had previously received a GnRH analogue. The median time from start of GnRH analogue to start of gender-affirming hormones was 11.3 months (range 2.2 to 42.0). 11 participants continued GnRH analogues after starting gender-affirming hormones.
		
		Average duration of treatment with a GnRH analogue not reported
		
		Comparison
		No comparator
			Critical Outcomes
			
			No critical outcomes assessed.
			
			Important outcomes
			
			Safety
			Of the 63 participants who received gender-affirming hormones:
			• No participants permanently discontinued gender-affirming hormones
			• 3 participants (5%) temporarily discontinued treatment:
			  ○ 2 transmales due to concomitant mental health comorbidities
			  ○ 1 transmale due to androgenic alopecia.
			  ○ No transfemale stopped treatment.
			The authors report that all patients eventually restarted gender-affirming hormones, although they do not report how long treatment was
				This study was appraised using the Newcastle-Ottawa tool for cohort studies.
				
				Domain 1: Selection domain
				1. b) somewhat representative
				2. c) no-non exposed cohort
				3. a) secure record*
				4. b) no
				Domain 2: Comparability
				1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders
				Domain 3: Outcome
				1. b) record linkage
				2. b) no – although follow-up time is reported for patients with more than 1 clinic visit, duration of treatment with gender-affirming hormones is not reported
				3. c) incomplete - missing data
				
				Overall quality is assessed as poor

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			stopped for, or what the effect of stopped treatment was. • No participants reported major complications • 12 participants (19%) had minor complications: ○ 7 transmales had severe acne (requiring isotretinoin) ○ 1 transmale had andogenic alopecia ○ 3 transmales had mild dyslipidaemia (levels not reported) ○ 1 transmale had significant mood swings ○ No transfemales had minor complications	Other comments: Mental health comorbidity was reported for all participants but not for the gender-affirming hormone cohort separately. Concomitant use of other medicines was not reported. Source of funding: No source of funding identified.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Klaver, Maartje, de Mutsert, Renee, van der Loos, Maria A T C et al. (2020) Hormonal Treatment and Cardiovascular Risk Profile in Transgender Adolescents. Pediatrics 145(3) **Study location** Single centre, Amsterdam, Netherlands	Participants were included if i) they had started GnRH analogue treatment before 18 years, ii) if whole body dual-energy radiograph absorptiometry was performed at least once during treatment (4 months before or after the start of GnRH analogues or gender-affirming hormones, or	Transfemales: Oestrogen (17-β oestradiol [E2]) orally, starting with 5 mcg/kg body weight per day, which was increased every 6 months until the maintenance dose of 2 mg per day was reached. Transmales: mixed testosterone esters (Sustanon), 25 mg/m^2 body surface area every 2 weeks intramuscularly,	**Critical Outcomes** No critical outcomes assessed. **Important outcomes** ***Safety*** Safety outcomes reported separately for transfemales and transmales. **For transfemales**, from the start of gender-affirming hormone treatment to age 22 years: • Mean BMI statistically significantly increased (mean change +1.9, 95% CI 0.6 to 3.2, p<0.005; mean BMI at	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Study type Retrospective chart review **Study aim** To examine the effects of treatment on changes in cardiovascular risk factors, including BMI, blood pressure, insulin sensitivity, and lipid levels. **Study dates** 1998-2015	within 1.5 years before or after the 22nd birthday), iii) if they were likely to have had at least 1 medical consultation in young adulthood. The study included 192 young people with dysphoria who met the above inclusion criteria: 71 transfemales and 121 transmales. Gender dysphoria was diagnosed according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition criteria. Mean age at the start of gender-affirming hormones was 16.4 years (SD 1.1) for transfemales and 16.9 years (SD 0.9) for transmales.	increased every 6 months to maintenance dose of 250 mg every 3 to 4 weeks. When GnRH analogues were started after the age of 16 years a different hormone starter dose was used (1 mg oestrogen daily and 75 mg testosterone weekly). Median (IQR) duration of GnRH analogue (monotherapy) was 2.1 years (1.0 to 2.7) in transfemales and 1.0 (0.5 to 2.9) for transmales.	22 years= 23.2, 95% CI 21.6 to 24.8). At age 22 years, 9.9% of the cohort were obese, compared with 3.0% in reference cisgender population[1]. • Mean systolic blood pressure (SBP) did not significantly change (mean change -3 mmHg, 95% CI -8 to 2; mean SBP at 22 years= 117 mmHg, 95% CI 113 to 122) • Mean diastolic blood pressure (DBP) statistically significantly increased (mean change +6 mmHg, 95% CI 3 to 10, p<0.001; mean DBP at 22 years= 75 mmHg, 95% CI 72 to 78) • Mean glucose level did not significantly change (mean change +0.1 mmol/L, 95% CI -0.1 to 0.2; mean glucose level at 22 years= 5.0 mmol/L, 95% CI 4.8 to 5.1) • Mean insulin level did not significantly change (mean change +2.7 mU/L, 95% CI -1.7 to 7.1; mean insulin level at 22 years= 5.0 mU/L (4.8 to 5.1) • Insulin resistance (mean Homeostatic Model Assessment of Insulin Resistance [HOMA-IR]) did not significantly change (mean change +0.7, 95% CI -0.2 to 1.5; mean HOMA-IR at 22 years 2.9, 95% CI 1.9 to 3.9) • Mean total cholesterol did not significantly change (mean change +0.1 mmol/L, 95% CI -0.2 to 0.4; mean total cholesterol at 22 years 4.1 mmol/L, 95% CI 3.8 to 4.4) • Mean HDL cholesterol did not significantly change (mean change +0.0 mmol/L, 95% CI -0.1 to 0.2; mean HDL cholesterol at 22 years 1.6 mmol/L, 95% CI 1.4 to 1.7) • Mean LDL cholesterol did not significantly change (mean change +0.0 mmol/L, 95%	of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. a) yes- follow-up from start of gender-affirming hormones to age 22 years, around 5 years 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor** Other comments: None Source of funding: No external funding

			CI -0.3 to 0.2; mean LDL cholesterol at 22 years 2.0 mmol/L, 95% CI 1.8 to 2.3) • Mean triglycerides statistically significantly increased (mean change +0.2 mmol/L, 95% CI 0.0 to 0.5, p<0.05; triglyceride level at 22 years 1.1 mmol/L, 95% CI 0.9 to 1.4) **For transmales**, from the start of gender-affirming hormone treatment to age 22 years: • Mean BMI statistically significantly increased (mean change +1.4, 95% CI 0.8 to 2.0, p<0.005; mean BMI at 22 years= 23.9, 95% CI 23.0 to 24.7). At age 22 years, 6.6% of the cohort were obese, compared with 2.2% in reference cisgender population[1]. • Mean systolic blood pressure (SBP) statistically significantly increased (mean change +5 mmHg, 95% CI 1 to 9; mean SBP at 22 years= 126 mmHg, 95% CI 122 to 130) • Mean diastolic blood pressure (DBP) statistically significantly increased (mean change +6 mmHg, 95% CI 4 to 9, p<0.001; mean DBP at 22 years= 74 mmHg, 95% CI 72 to 77) • Mean glucose level did not significantly change (mean change 0.0 mmol/L, 95% CI -0.2 to 0.2; mean glucose level at 22 years= 4.8 mmol/L, 95% CI 4.7 to 5.0) • Mean insulin level statistically significantly decreased (mean change -2.1 mU/L, 95% CI -3.9 to -0.3, p<0.05; mean insulin level at 22 years= 8.6 mU/L (6.9 to 10.2) • Insulin resistance (mean Homeostatic Model Assessment of Insulin Resistance [HOMA-IR]) statistically significantly	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			decreased (mean change -0.5, 95% CI -1.0 to -0.1, p<0.05; mean HOMA-IR at 22 years 1.8, 95% CI 1.4 to 2.2) • Mean total cholesterol statistically significantly increased (mean change +0.4 mmol/L, 95% CI 0.2 to 0.6, p<0.001; mean total cholesterol at 22 years 4.6 mmol/L, 95% CI 4.3 to 4.8) • Mean HDL cholesterol statistically significantly decreased (mean change -0.3 mmol/L, 95% CI -0.4 to -0.2, p<0.001; mean HDL cholesterol at 22 years 1.3 mmol/L, 95% CI 1.2 to 1.3) • Mean LDL cholesterol statistically significantly increased (mean change +0.4 mmol/L, 95% CI 0.2 to 0.6, p<0.001; mean LDL cholesterol at 22 years 2.6 mmol/L, 95% CI 2.4 to 2.8) • Mean triglycerides statistically significantly increased (mean change +0.5 mmol/L, 95% CI 0.3 to 0.7, p<0.001; triglyceride level at 22 years 1.3 mmol/L, 95% CI 1.1 to 1.5)	

[1] Reference population taken from Fredriks et al. (2000)

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Klink D, Caris M, Heijboer A et al. (2015) Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. The Journal of Clinical Endocrinology and Metabolism 100(2): e270-5 **Study location** Single centre, Amsterdam, Netherlands **Study type** Retrospective longitudinal study **Study aim** To assess peak bone mass in young adults with gender dysphoria who had received GnRH analogues and gender-affirming hormones during their pubertal years. **Study dates**	34 young people with gender dysphoria who received GnRH analogues, gender-affirming hormones and gonadectomy. The study included 15 transfemales and 19 transmales; mean age at start of gender-affirming hormones was 16.6 years (SD 1.4) and 16.4 years (SD 2.3) respectively. Participants were required to meet the DSM-IV-TR criteria for gender identity disorder of adolescence. Participants were included if they had undergone gonadectomy between June 1998 and August 2012, and they were at least 21 years old when they had the surgery. Bone mineral density data were also required at the start of GnRH analogue, gender-affirming hormones and at the age of 22 years. No concomitant treatments were reported.	**Intervention** Transfemales - oral 17-β oestradiol (incremental dosing) Transmales – IM testosterone (Sustanon 250 mg/ml; incremental dosing) Median duration of treatment with gender-affirming hormones for transfemales was 5.8 years (range 3.0 to 8.0) and for transmales was 5.4 years (range 2.8 to 7.8). The GnRH analogue was SC triptorelin 3.75 mg every 4 weeks. No details of gonadectomy reported. **Comparison** No comparison group. Comparison over time reported.	**Critical outcomes** No critical outcomes reported **Important outcomes** **Safety** ***Bone density: lumbar spine*** **Lumbar spine bone mineral apparent density (BMAD)** Change from starting gender-affirming hormones to age 22 years in transfemales-Mean (SD); g/m^3 • Start of gender-affirming hormones: 0.22 (0.02) • Age 22 years: 0.23 (0.03) • p=0.003 z-score (range) • Start of gender-affirming hormones: -0.90 (0.80) • Age 22 years: -0.78 (1.03) • No statistically significant difference Change from starting gender-affirming hormones to age 22 years in transmales-Mean (SD); g/m^3 • Start of gender-affirming hormones: 0.24 (0.02) • Age 22 years: 0.25 (0.28 • p=0.001 z-score (SD) • Start of gender-affirming hormones: -0.50 (0.81) • Age 22 years: -0.033 (0.95) • p=0.002	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. a) yes – mean duration of gender-affirming hormone treatment was 5.8 and 5.4 years. 3. c) follow-up rate variable across timepoints and no description of those lost **Overall quality is assessed as poor** Other comments: Within person comparison. Small numbers of participants in each subgroup. No

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Gonadectomy took place between June 1998 and August 2012	At the start of gender-affirming hormone treatment, in the transfemale subgroup the median Tanner P was 4 (IQR 2) and the median Tanner G was 12 (IQR 11). In the transmale subgroup the median Tanner B was 5 (IQR 2) and the median Tanner P was 5 (IQR 0).		**Lumbar spine bone mineral density (BMD)** Change from starting gender-affirming hormones to age 22 years in transfemales-Mean (SD); g/m^2 • Start of gender-affirming hormones: 0.84 (0.11) • Age 22 years: 0.93 (0.10) • p<0.001 z-score (range) • Start of gender-affirming hormones: -1.01 (0.98) • Age 22 years: -1.36 (0.83) • No statistically significant difference Change from starting gender-affirming hormones to age 22 years in transmales-Mean (SD); g/m^2 • Start of gender-affirming hormones: 0.91 (0.10) • Age 22 years: 0.99 (0.13) • P<0.001 z-score (range) • Start of gender-affirming hormones: -0.72 (0.99) • Age 22 years: -0.33 (1.12) • No statistically significant difference ***Bone density: femoral region, nondominant side*** **Femoral region, nondominant side BMAD** Change from starting gender-affirming hormones to age 22 years in transfemales-Mean (SD); g/m^3 • Start of gender-affirming hormones: 0.26 (0.04) • Age 22 years: 0.28 (0.05)	concomitant treatments or comorbidities were reported. Source of funding: None disclosed

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			• No statistically significant difference z-score (SD) • Start of gender-affirming hormones: -1.57 (1.74) • Age 22 years: Not reported • No statistical analysis reported Change from starting gender-affirming hormones to age 22 years in transmales-Mean (SD); g/m^3 • Start of gender-affirming hormones: 0.31 (0.04) • Age 22 years: 0.33 (0.05) • p=0.010 z-score (SD) • Start of gender-affirming hormones: -0.28 (0.74) • Age 22 years: Not reported • No statistical analysis reported **Femoral region, nondominant side BMD** Change from starting gender-affirming hormones to age 22 years in transfemales-Mean (SD); g/m^2 • Start of gender-affirming hormones: 0.87 (0.08) • Age 22 years: 0.94 (0.11) • P=0.009 z-score (SD) • Start of gender-affirming hormones: -0.95 (0.63) • Age 22 years: -0.69 (0.74) • No statistically significant difference Change from starting gender-affirming hormones to age 22 years in transmales-Mean (SD); g/m^2 • Start of gender-affirming hormones: 0.88 (0.09) • Age 22 years: 0.95 (0.10)	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			• P<0.001 z-score (SD) • Start of gender-affirming hormones: -0.35 (0.79) • Age 22 years: -0.35 (0.74) • p=0.006	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Kuper, Laura E, Stewart, Sunita, Preston, Stephanie et al. (2020) Body Dissatisfaction and Mental Health Outcomes of Youth on Gender-Affirming Hormone Therapy. Pediatrics 145(4) **Study location** Single centre, Texas, USA **Study type** Prospective longitudinal study **Study aim** To: • explore how baseline body dissatisfaction, depression, and anxiety symptoms vary by gender,	148 children and adolescents with gender dysphoria, n=148, of whom: • 25 received puberty suppression only • 93 received gender-affirming hormone therapy only • 30 received both Results for treatments reported separately. Mean age at initial assessment was 15.4 years (range 9 to 18). Mean age at start of gender-affirming hormone therapy was 16.2 years (range 13.2 to 18.6). All participants met the Diagnostic and Statistical	Hormone therapy, guided by Endocrine Society Clinical Practice Guidelines Follow-up at least 18 months from initial assessment at the clinic. Mean duration of gender-affirming hormone therapy before follow-up was 10.9 months (range 1 to 18; SD 3.3)	**Critical Outcomes** ***Impact on mental health*** Mean depression score, assessed using the Quick Inventory of Depressive Symptoms (QIDS), self-reported was 9.6 (SD 5.0) at baseline and 7.4 (SD 4.5) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean depression score, assessed using the QIDS, clinician-reported was 5.9 (SD 4.1) at baseline and 6.0 (SD 3.8) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean anxiety score, assessed using the Screen for Child Anxiety Related Emotional Disorders (SCARED) questionnaire was 32.6 (SD 16.3) at baseline and 28.4 (SD 15.9) at	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. d) assessors not blinded to treatment 2. a) yes – follow-up at least 18 months from initial assessment. Mean duration of gender-affirming hormone

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
age at initial assessment, and Tanner stage at first medical visit • examine how body dissatisfaction, depression, and anxiety symptoms change over the first year of gender-affirming hormone treatment • explore how any changes vary by affirmed gender, Tanner stage, age, type of treatment, months on gender-affirming hormone therapy, mental health treatment received, and whether chest surgery was also obtained (among transmales). **Study dates** Initial participant assessments took place between August 2014 and March 2018.	Manual of Mental Disorders, Fifth Edition criteria for gender dysphoria. Specific inclusion and exclusion criteria for the study are not reported. It would appear that all children and adolescents eligible for gender-affirming hormones were considered eligible for the study. The authors state that before initial assessment with a psychologist, psychiatrist, and/or clinical therapist, parents completed a phone intake survey. Around one-third of families did not follow-up after the phone intake.		follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean panic score, assessed using specific questions from the SCARED questionnaire was 8.1 (SD 6.3) at baseline and 7.1 (SD 6.5) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean generalised anxiety score, assessed using specific questions from the SCARED questionnaire was 10.0 (SD 5.1) at baseline and 8.8 (SD 6.5) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean social anxiety score, assessed using specific questions from the SCARED questionnaire was 8.5 (SD 4.1) at baseline and 7.7 (SD 4.2) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean separation anxiety score, assessed using specific questions from the SCARED	treatment was 10.9 months. 3. c) patient numbers vary by outcome with no explanation **Overall quality is assessed as poor** Other comments: None Source of funding: Supported by Children's Health. The Research Electronic Data Capture database was funded by the Clinical and Translational Science Awards program

			questionnaire was 3.5 (SD 3.0) at baseline and 3.1 (SD 2.5) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. Mean school avoidance score, assessed using specific questions from the SCARED questionnaire was 2.6 (SD 2.1) at baseline and 2.0 (SD 2.0) at follow-up. The authors did not present statistical analysis for the sub-group of participants receiving gender-affirming hormones and it is unclear whether the change in score was statistically significant. The authors also reported results separately for transfemales and transmales: **Transfemales** No statistical analyses were reported for this sub-group and it is unclear whether any changes in score were statistically significant. • Mean depression symptoms, assessed using the QIDS, self-reported was 7.5 (SD 4.9) at baseline and 6.6 (SD 4.4) at follow-up. • Mean depression symptoms, assessed using the QIDS, clinician-reported was 4.2 (SD 3.2) at baseline and 5.4 (SD 3.4) at follow-up. • Mean anxiety symptoms, assessed using the SCARED questionnaire was 26.4 (SD 14.2) at baseline and 24.3 (SD 15.4) at follow-up.	

			• Mean panic symptoms, assessed using specific questions from the SCARED questionnaire was 5.7 (SD 4.9) at baseline and 5.1 (SD 4.9) at follow-up. • Mean generalised anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 8.6 (SD 5.1) at baseline and 8.0 (SD 5.1) at follow-up. • Mean social anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 7.1 (SD 3.9) at baseline and 6.8 (SD 4.4) at follow-up. • Mean separation anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 3.4 (SD 3.3) at baseline and 2.7 (SD 2.3) at follow-up. • Mean school avoidance symptoms, assessed using specific questions from the SCARED questionnaire was 1.8 (SD 1.7) at baseline and 1.9 (SD 2.1) at follow-up. **Transmales** No statistical analyses were reported for this sub-group and it is unclear whether any changes in score were statistically significant. • Mean depression symptoms, assessed using the QIDS, self-reported was 10.4 (SD 5.0) at baseline and 7.5 (SD 4.5) at follow-up. • Mean depression symptoms, assessed using the QIDS, clinician-reported was 6.7 (SD 4.4) at baseline and 6.2 (SD 4.1) at follow-up. • Mean anxiety symptoms, assessed using the SCARED questionnaire was 35.4 (SD	

			16.5) at baseline and 29.8 (SD 15.5) at follow-up. • Mean panic symptoms, assessed using specific questions from the SCARED questionnaire was 9.3 (SD 6.5) at baseline and 7.9 (SD 6.5) at follow-up. • Mean generalised anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 10.4 (SD 5.0) at baseline and 9.0 (SD 5.1) at follow-up. • Mean social anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 8.5 (SD 4.0) at baseline and 7.8 (SD 4.1) at follow-up. • Mean separation anxiety symptoms, assessed using specific questions from the SCARED questionnaire was 4.2 (SD 3.4) at baseline and 3.4 (SD 2.6) at follow-up. • Mean school avoidance symptoms, assessed using specific questions from the SCARED questionnaire was 2.6 (SD 2.1) at baseline and 2.0 (SD 2.0) at follow-up. No difference in impact on mental health found by Tanner age. Numerical results, statistical analysis and information on specific outcomes not reported. It is unclear from the paper whether Tanner age is at initial assessment, start of GnRH analogues, start of gender-affirming hormones, or another timepoint. **Important Outcomes** ***Impact on body image***	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			Mean Body Image Scale (BIS) score was 70.7 (SD 15.2) at baseline and 51.4 (SD 18.3) at follow-up. The authors do not present statistical analysis for this population and it is unclear whether the change in score was statistically significant. The authors also reported body image results separately for transfemales and transmales. No statistical analyses were reported for this sub-groups and it is unclear whether changes in score were statistically significant. • In transfemales, BIS score was 67.5 (SD 19.5) at baseline and 49.0 (SD 21.6) at follow-up. • In transmales, BIS score was 71.1 (SD 13.4) at baseline and 52.9 (SD 16.8) at follow-up. No difference in body image score found by Tanner age. Numerical results, statistical analysis and information on specific outcomes not reported. It is unclear from the paper whether Tanner age is at initial assessment, start of GnRH analogues, start of gender-affirming hormones, or another timepoint. *No other critical or important outcomes reported*	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Study dates Lopez de Lara, D., Perez Rodriguez, O., Cuellar Flores, I. et al. (2020) Psychosocial assessment in transgender adolescents. Anales de Pediatria **Study location** Single centre in Madrid, Spain **Study type** Prospective analytical study **Study aim** To assess the psychosocial status of patients seeking care in the paediatric endocrinology clinic for gender dysphoria, and the impact on psychosocial status of gender-affirming hormone therapy at 12 months of treatment **Study dates** Not reported	23 adolescents with gender dysphoria; 16 transmale and 7 transfemale. Participants were required to be at a stage of pubertal development of Tanner 2 or higher. People with mental health comorbidity that could affect the experience of gender dysphoria were excluded. Mean age at baseline was 16 years (range 14 to 18). 30 cisgender controls, matched for age, ethnicity, and socioeconomic status	Gender-affirming hormones- • Oral oestradiol • Intramuscular testosterone Participants had previously received gonadotropin-releasing hormone (GnRH) analogues in the intermediate pubertal stages (Tanner 2---3).	**Critical Outcomes** ***Impact on gender dysphoria*** Following gender-affirming hormones for 12 months, mean (±SD) Utrecht Gender Dysphoria Scale (UGDS) score statistically significantly improved, from 57.1 (±4.1) at baseline to 14.7 (±3.2; p<0.001) ***Impact on mental health*** Mean depression score statistically significantly improved following treatment with gender-affirming hormones. Mean Beck Depression Inventory II (BDI-II) score (±SD) reduced from 19.3 points (±5.5) at baseline to 9.7 points (±3.9) at 12 months (p<0.001). Mean anxiety scores statistically significantly improved following treatment with gender-affirming hormones. Mean (±SD) State-Trait Anxiety Inventory (STAI) State subscale score improved from 33.3 points (±9.1) at baseline to 16.8 points (±8.1) at 12 months (p<0.001). Mean (±SD) State-Trait Anxiety Inventory (STAI) Trait subscale score improved from 33.0 points (±7.2) at baseline to 18.5 points (±8.4) at 12 months (p<0.001). **Important Outcomes** ***Psychosocial Impact*** There was not change in family functioning, measured using the Family APGAR test, from baseline (17.9 points) to 1 year after starting	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. Not applicable – although a control group is reported on, people in this group did not have gender dysphoria. 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. Not applicable – although a control group is reported on, people in this group did not have gender dysphoria. **Domain 3: Outcome** 1. d) assessors not blinded to treatment 2. a) yes – 12 months treatment with gender-affirming hormones 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor**

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			gender-affirming hormones (18.0 points; no statistical analysis reported). Results from the Strengths and Difficulties Questionnaire, Spanish Version (SDQ-Cas) showed statistically significant improvements from baseline (14.7 points; SD±3.3) to 12 months after gender-affirming hormones (10.3 points; SD±2.9; p<0.001) *No other critical or important outcomes reported*	Other comments: None Source of funding: Not reported

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Stoffers, Iris E; de Vries, Martine C; Hannema, Sabine E (2019) Physical changes, laboratory parameters, and bone mineral density during testosterone treatment in adolescents with gender dysphoria. The journal of sexual medicine 16(9): 1459-1468 **Study location** Single centre, Leiden, Netherlands **Study type** Retrospective chart review **Study aim** To report changes in height, BMI, blood pressure, laboratory parameters and bone density. **Study dates** November 2010 to August 2018	62 transmales with gender dysphoria. participants were required to have been receiving testosterone therapy for at least 6 months. Further inclusion or exclusion criteria not reported. Gender dysphoria was diagnosed according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition criteria. .	Testosterone intramuscular injection (Sustanon 250 mg). Dose escalated every 6 months up to the standard adult dose of 125 mg every 2 weeks or 250 mg every 3-4 weeks. A more rapid dose escalation was using in patients who started GnRH analogue treatment at 16 years or older. Median age at start of testosterone treatment was 17.2 years (range 14.9 to 18.4) Median duration of testosterone treatment was 12 months (range 5 to 33) Median duration of GnRH analogue treatment was 8 months (range 3 to 39)	**Critical Outcomes** No critical outcomes assessed. **Important outcomes** ***Safety*** ***Bone mineral density (BMD): lumbar spine*** There was no statistically significant difference in lumber spine bone mineral density (BMD) from start of testosterone treatment to any timepoint, up to 24 months follow-up. Mean (±SD), g/cm^2: • Start of testosterone: 0.90 (±0.11) • 6 months: 0.94 (±0.10) • 12 months: 0.95 (±0.09) • 24 months: 0.95 (±0.11) z-score (±SD): • Start of testosterone: -0.81 (±1.02) • 6 months: -0.67 (±0.95) • 12 months: -0.66 (±0.81) • 24 months: -0.74 (±1.17) ***Bone mineral density (BMD): femoral neck (hip)*** There was no statistically significant difference in right or left femoral neck (hip) bone mineral density (BMD) from start of	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. a) yes – mean duration of gender-affirming hormone treatment was 5.8 and 5.4 years. 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor** Other comments: None Source of funding: None

			testosterone treatment to any timepoint, up to 24 months follow-up.
			Right
			Mean (±SD), g/cm^2:
			• Start of testosterone: 0.77 (±0.08)
			• 6 months: 0.84 (±0.11)
			• 12 months: 0.82 (±0.08)
			• 24 months: 0.85 (±0.11)
			z-score (±SD):
			• Start of testosterone: -0.97 (0.79)
			• 6 months: -0.54 (±0.96)
			• 12 months: -0.80 (±0.69)
			• 24 months: -0.31 (±0.84)
			Left
			Mean (±SD), g/cm^2:
			• Start of testosterone: 0.76 (±0.09)
			• 6 months: 0.83 (±0.12)
			• 12 months: 0.81 (±0.08)
			• 24 months: 0.86 (±0.09)
			z-score (±SD):
			• Start of testosterone: -1.07 (0.85)
			• 6 months: -0.62 (±1.12)
			• 12 months: -0.93 (±0.63)
			• 24 months: -0.20 (±0.70)
			
			Other safety-related outcomes
			• Alkaline phosphatase: statistically significant increases observed from start of testosterone treatment to 6 months and 12 months ($p<0.001$), although difference at 24 months was not statistically significant. Median (IQR), U/L
			  ○ Start of testosterone: 102 (78 to 136)
			  ○ 6 months: 115 (102 to 147)
			  ○ 12 months: 112 (88 to 143)
			  ○ 24 months: 81 (range 69 to 98)
			• Creatinine: statistically significant increases observed from start of	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			testosterone treatment to 6, 12 and 24 months (p<0.001). Mean (±SD), umol/L ○ Start of testosterone: 62 (±7) ○ 6 months: 70 (±9) ○ 12 months: 74 (±10) ○ 24 months: 81 (±10) There was no statistically significant change from start of testosterone treatment in: • HbA1c • Aspartate aminotransferase (AST) • Alanine aminotransferase (ALT) • Gamma-glutamyl transferase • Urea Numerical results, follow-up duration and further details of statistical analysis not reported.	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Full citation Vlot MC, Klink DT, den Heijer M et al. (2017) Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone 95: 11-19 **Study location** Single centre, Amsterdam, Netherlands **Study type** Retrospective chart review **Study aim** To investigate the impact of GnRH analogues and gender-affirming hormones on bone mineral apparent density (BMAD) in transgender adolescents. The study also report on levels of bone turnover markers, although the authors concluded that the	70 adolescents with gender dysphoria (42 transmales and 28 transfemales). Median age (range) at the start of gender-affirming hormones was 16.3 years (15.9 to 19.5) for transmales and 16.0 years (14.0 to 18.9) for transfemales. Participants were included if they had a diagnosis of gender dysphoria according to DSM-IV-TR criteria who received GnRH analogues and then gender-affirming hormones. No concomitant treatments were reported. The study categorised participants into a young and old pubertal group, based on their bone age. The young transmales had a bone age of <14 years and the old transmales had a bone age of ≥14 years. The young transfemales	Transfemales: Oestradiol oral Dose escalated every 6 months until standard adult dose of 2 mg daily was reached Transmales: Testosterone intramuscular injection (Sustanon 250 mg). Dose escalated every 6 months up to the standard adult dose of 250 mg every 4 weeks or 250 mg every 3-4 weeks. All participants previously received a GnRH analogue (triptorelin 3.75 mg subcutaneously every 4 weeks) Median duration of GnRH analogue therapy not reported.	**Critical outcomes** No critical outcomes reported **Important outcomes** ***Bone density: lumbar spine*** **Lumbar spine bone mineral apparent density (BMAD)** Transfemales (bone age <15 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones (C0): 0.20 (0.18 to 0.24) • 24-month follow-up (C24): 0.22 (0.19 to 0.27) • Statistically significant increase (p≤0.01) z-score (range) • Start of gender-affirming hormones (C0): -1.52 (-2.36 to 0.42) • 24-month follow-up (C24): • Statistically significant increase (p≤0.05) Transfemales (bone age ≥15 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones: 0.22 (0.19 to 0.24) • 24-months: 0.23 (0.21 to 0.26) • Statistically significant increase (p≤0.05) z-score (range) • Start of gender-affirming hormones: -1.15 (-2.21 to 0.08) • 24-months: -0.66 (-1.66 to 0.54)	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain** 1. b) somewhat representative 2. c) no-non exposed cohort 3. a) secure record* 4. b) no **Domain 2: Comparability** 1. c) cohorts are not comparable on the basis of the design or analysis controlled for confounders **Domain 3: Outcome** 1. b) record linkage 2. a) yes- 24 month follow-up 3. a) complete follow up - all subjects accounted for **Overall quality is assessed as poor.** Other comments: None Source of funding: grant from Abbott diagnostics

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
added value of these seems to be limited. **Study dates** Participants started gender-affirming therapy between 2001 and 2011	group had a bone age of <15 years and the old transfemales group ≥15 years.		Statistically significant increase (p≤0.05) Transmales (bone age <14 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones: 0.23 (0.19 to 0.28) • 24-months: 0.25 (0.22 to 0.28) • Statistically significant increase (p≤0.01) z-score (range) • Start of gender-affirming hormones: -0.84 (-2.2 to 0.87) • 24-months: -0.15 (-1.38 to 0.94) Statistically significant increase (p≤0.01) Transmales (bone age ≥14 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones: 0.24 (0.20 to 0.28) • 24-months: 0.25 (0.21 to 0.30) • Statistically significant increase (p≤0.01) z-score (range) • Start of gender-affirming hormones: -0.29 (-2.28 to 0.90) • 24-months: -0.06 (-1.75 to 1.61) Statistically significant increase (p≤0.01) ***Bone density: femoral neck*** **Femoral neck BMAD** Transfemales (bone age <15 years), change from starting gender-affirming hormones to 24 months follow-up.	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			Median (range), g/m^3 • Start of gender-affirming hormones: 0.27 (0.20 to 0.33) • 24-months: 0.27 (0.20 to 0.36) • No statistically significant change z-score (range) • Start of gender-affirming hormones: -1.32 (-3.39 to 0.21) • 24-months: -1.30 (-3.51 to 0.92) • No statistically significant change Transfemales (bone age ≥15 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones: 0.30 (0.26 to 0.34) • 24-months: 0.29 (0.24 to 0.38) • No statistically significant change z-score (range) • Start of gender-affirming hormones: -0.36 (-1.50 to 0.46) • 24-months: -0.56 (-2.17 to 1.29) • No statistically significant change Transmales (bone age <14 years), change from starting gender-affirming hormones to 24 months follow-up. Median (range), g/m^3 • Start of gender-affirming hormones: 0.30 (0.22 to 0.35) • 24-months: 0.33 (0.23 to 0.37) • Statistically significant increase (p≤0.01) z-score (range) • Start of gender-affirming hormones: -0.37 (-2.28 to 0.47) • 24-months: -0.37 (-2.03 to 0.85)	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			• Statistically significant increase (p≤0.01) Transmales (bone age ≥14 years), change from starting gender-affirming hormones to 24 months follow-up. • Start of gender-affirming hormones: 0.30 (0.23 to 0.41) • 24-months: 0.32 (0.23 to 0.41) • Statistically significant increase (p≤0.01) z-score (range) • Start of gender-affirming hormones: -0.27 ((-1.91 to 1.29) • 24-months: 0.02 (-2.1 to 1.35) • Statistically significant increase (p≤0.05)	

Appendix F Quality appraisal checklists

Newcastle-Ottawa Quality Assessment Form for Cohort Studies

Note: A study can be given a maximum of one star for each numbered item within the Selection and Outcome categories. A maximum of two stars can be given for Comparability.

Selection

1) Representativeness of the exposed cohort
    a) Truly representative (one star)
    b) Somewhat representative (one star)
    c) Selected group
    d) No description of the derivation of the cohort
2) Selection of the non-exposed cohort
    a) Drawn from the same community as the exposed cohort (one star)
    b) Drawn from a different source
    c) No description of the derivation of the non exposed cohort
3) Ascertainment of exposure
    a) Secure record (e.g., surgical record) (one star)
    b) Structured interview (one star)
    c) Written self report
    d) No description
    e) Other
4) Demonstration that outcome of interest was not present at start of study
    a) Yes (one star)
    b) No

Comparability

1) Comparability of cohorts on the basis of the design or analysis controlled for confounders
    a) The study controls for age, sex and marital status (one star)
    b) Study controls for other factors (list) _____ (one star)
    c) Cohorts are not comparable on the basis of the design or analysis controlled for confounders

Outcome

1) Assessment of outcome
    a) Independent blind assessment (one star)
    b) Record linkage (one star)
    c) Self report
    d) No description
    e) Other
2) Was follow-up long enough for outcomes to occur
    a) Yes (one star)
    b) No
    Indicate the median duration of follow-up and a brief rationale for the assessment above:_____
3) Adequacy of follow-up of cohorts
    a) Complete follow up- all subject accounted for (one star)

b) Subjects lost to follow up unlikely to introduce bias- number lost less than or equal to 20% or description of those lost suggested no different from those followed. (one star)
c) Follow up rate less than 80% and no description of those lost
d) No statement

Thresholds for converting the Newcastle-Ottawa scales to AHRQ standards (good, fair, and poor):

Good quality: 3 or 4 stars in selection domain AND 1 or 2 stars in comparability domain AND 2 or 3 stars in outcome/exposure domain

Fair quality: 2 stars in selection domain AND 1 or 2 stars in comparability domain AND 2 or 3 stars in outcome/exposure domain

Poor quality: 0 or 1 star in selection domain OR 0 stars in comparability domain OR 0 or 1 stars in outcome/exposure domain

Appendix G Grade profiles

Table 2: Question 1: For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? - Gender dysphoria

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on gender dysphoria (1 uncontrolled, prospective observational study)									
Change from baseline in mean gender dysphoria score, measured using the UGDS (duration of treatment 12 months). Higher scores indicate greater gender dysphoria.									
1 cohort study Lopez de Lara et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 57.1 (SD 4.1) T1 (12 months) = 14.7 (SD 3.2) Statistically significant improvement, p<0.001	Critical	VERY LOW

Abbreviations: p: p-value; SD: standard deviation; UGDS: Utrecht Gender Dysphoria Scale

1 Downgraded 1 level - the cohort study by Lopez de Lara et al. 2020 was assessed at high risk of bias (poor quality overall; lack of blinding and no control group)

Table 3: Question 1: For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Mental health

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on mental health (3 uncontrolled, prospective observational studies and 2 uncontrolled, retrospective observational studies)									
Change from baseline in mean depression score, measured using the BDI-II (duration of treatment 12 months). Higher scores indicate more severe depression.									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 cohort study Lopez de Lara et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 19.3 (SD 5.5) T1 (12 months) = 9.7 (SD 3.9) Statistically significant improvement, p<0.001	Critical	VERY LOW
Change from baseline in mean depression score, measured using the CESD-R (approximately 12-month follow-up). Higher scores indicate more severe depression.									
1 cohort study Achille et al. 2020	Serious limitations[2]	Serious indirectness[3]	No serious inconsistency	Not calculable	N=50	None	Wave 1 (baseline) = 21.4 Wave 3 (approx. 12 months) = 13.9 Statistically significant improvement (p<0.001)	Critical	VERY LOW
Change from baseline in depression score, measured using the Patient Health Questionnaire Modified for Teens (PHQ 9_Modified for Teens) (approximately 12-month follow-up). Higher scores indicate more severe depression.									
1 cohort study Achille et al. 2020	Serious limitations[2]	Serious indirectness[3]	No serious inconsistency	Not calculable	N=50	None	Statistically significant reductions in mean score, p<0.001 Results presented diagrammatically, numerical results for mean score not reported	Critical	VERY LOW
Change from baseline in depression symptoms, measured using the Quick Inventory of Depressive Symptoms (QIDS), self-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe depression.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=105	None	Baseline = 9.6 (SD 5.0) Follow-up = 7.4 (SD 4.5) No statistical analysis reported for the sub-group of participants receiving gender-affirming hormones	Critical	VERY LOW
Change from baseline in depression symptoms, measured using the Quick Inventory of Depressive Symptoms (QIDS), clinician-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe depression.									
1 cohort study	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=106	None	Baseline = 5.9 (SD 4.1) Follow-up = 6.0 (SD 3.8)	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Kuper et al. 2020							No statistical analysis reported for the sub-group of participants who received gender-affirming hormones		
Need for treatment due to depression, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 54% (28/52) During real life phase 15% (8/52) Statistically significant reduction (p<0.001)	Critical	VERY LOW
Change from baseline in anxiety score, measured using the STAI-State subscale (duration of treatment 12 months). Higher scores indicate more severe anxiety.									
1 cohort study Lopez de Lara et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 33.3 (SD 9.1) T1 (12 months) = 16.8 (SD 8.1) Statistically significant improvement, p<0.001	Critical	VERY LOW
Change from baseline in anxiety score, measured using the STAI-Trait subscale (duration of treatment 12 months). Higher scores indicate more severe anxiety.									
1 cohort study Lopez de Lara et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 33.0 (SD 7.2) T1 (12 months) = 18.5 (SD 8.4) Statistically significant improvement, p<0.001	Critical	VERY LOW
Change from baseline in anxiety symptoms, measured using the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe anxiety.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=80	None	Baseline = 32.6 (SD 16.3) Follow-up = 28.4 (SD 15.9) No statistical analysis reported for the sub-group of participants	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							who received gender-affirming hormones		
Change from baseline in panic symptoms, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=82	None	Baseline = 8.1 (SD 6.3) Follow-up = 7.1 (SD 6.5) No statistical analysis reported for the sub-group of participants who received gender-affirming hormones	Critical	VERY LOW
Change from baseline in generalised anxiety symptoms, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=82	None	Baseline = 10.0 (SD 5.1) Follow-up = 8.8 (SD 5.0) No statistical analysis reported for the sub-group of participants who received gender-affirming hormones	Critical	VERY LOW
Change from baseline in social anxiety symptoms, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=82	None	Baseline = 8.5 (SD 4.1) Follow-up = 7.7 (SD 4.2) No statistical analysis reported for the sub-group of participants who received gender-affirming hormones	Critical	VERY LOW
Change from baseline in separation anxiety symptoms, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=81	None	Baseline = 3.5 (SD 3.0) Follow-up = 3.1 (SD 2.5) No statistical analysis reported for the sub-group of participants	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							who received gender-affirming hormones		
Change from baseline in school avoidance, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=80	None	Baseline = 2.6 (SD 2.1) Follow-up = 2.0 (SD 2.0) No statistical analysis reported for the sub-group of participants who received gender-affirming hormones	Critical	VERY LOW
Need for treatment due to anxiety, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 48% (25/52) During real life phase 15% (8/52) Statistically significant reduction (p<0.001)	Critical	VERY LOW
Change from baseline in adjusted mean suicidality score, measured using the ASQ instrument (mean treatment duration 349 days). Higher scores indicate a greater degree of suicidality.									
1 cohort study Allen et al. 2019	Serious limitations[5]	No serious indirectness	No serious inconsistency	Not calculable	N=39	None	T0 (baseline) = 1.11 (SE 0.22) T1 (final assessment) = 0.27 (SE 0.12) Statistically significant improvement in score from T0 to T1, p<0.001	Critical	VERY LOW
Change from baseline in percentage of participants with suicidal ideation, measured using the additional questions from the PHQ 9_Modified for Teens (approximately 12-month follow-up)									
1 cohort study Achille et al. 2020	Serious limitations[2]	Serious indirectness[3]	No serious inconsistency	Not calculable	N=50	None	Wave 1 (baseline) = 10% (5/50) Wave 3 (approx. 12 months) = 6% (3/50)	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							No statistical analysis reported		
Change from baseline in suicidal ideation (passive), information on which was collected by clinician, exact methods / tools not reported (mean duration of gender-affirming hormone treatment was 10.9 months)									
1 cohort study Kuper et al. 2020	Serious limitations[4]	Serious indirectness [6]	No serious inconsistency	Not calculable	N=130	None	Lifetime = 81% (105 people) 1 month before initial assessment = 25% (33 people) Follow-up period = 38% (51 people) No statistical analysis reported	Critical	VERY LOW
Change from baseline in suicide attempts, information on which was collected by clinician, exact methods / tools not reported (mean duration of gender-affirming hormone treatment was 10.9 months)									
1 cohort study Kuper et al. 2020	Serious limitations[4]	Serious indirectness[6]	No serious inconsistency	Not calculable	N=130	None	Lifetime = 15% (20 people) 3 months before initial assessment = 2% (3 people) Follow-up period = 5% (6 people) No statistical analysis reported	Critical	VERY LOW
Change from baseline in non-suicidal self-injury, information on which was collected by clinician, exact methods / tools not reported (mean duration of gender-affirming hormone treatment was 10.9 months)									
1 cohort study Kuper et al. 2020	Serious limitations[4]	Serious indirectness[6]	No serious inconsistency	Not calculable	N=130	None	Lifetime = 52% (68 people) 3 months before initial assessment = 10% (13 people) Follow-up period = 17% (23 people) No statistical analysis reported	Critical	VERY LOW
Need for treatment due to suicidality / self-harm, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 35% (18/52) During real life phase	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							4% (2/52) Statistically significant reduction (p<0.001)		
Need for mental health treatment, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 50% (26/52) During real life phase 46% (24/51) No statistically significant difference (p= 0.77)	Critical	VERY LOW
Need for treatment due to conduct problems / antisocial, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 14% (7/52) During real life phase 6% (3/52) No statistically significant difference (p= 0.18)	Critical	VERY LOW
Need for treatment due to psychotic symptoms or psychosis, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 2% (1/52) During real life phase 4% (2/52) No statistically significant difference (p= 0.56)	Critical	VERY LOW
Need for treatment due to substance abuse, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 4% (2/52) During real life phase 2% (1/52) No statistically significant difference (p= 0.56)	Critical	VERY LOW
Need for treatment due to autism, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 12% (6/52) During real life phase 6% (3/52) No statistically significant difference (p= 0.30)	Critical	VERY LOW
Need for treatment due to ADHD, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 10% (5/52) During real life phase 2% (1/52) No statistically significant difference (p= 0.09)	Critical	VERY LOW
Need for treatment due to eating disorder, during and before gender identity assessment, and during real life phase (approximately 12 months follow-up)									
1 cohort study Kaltiala et al. 2020	Serious limitations[7]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During and before gender identity assessment 2% (1/52)	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							During real life phase 2% (1/52) No statistically significant difference (p=1.0)		

Abbreviations: ADHD: attention deficit hyperactivity disorder; ASQ: Ask Suicide-Screening Questions; CESD-R: Center for Epidemiologic Studies Depression Scale; BDI-II: Beck Depression Inventory II (BDI-II); p: p-value; PHQ 9_Modified for Teens: Patient Health Questionnaire Modified for Teens; SCARED: Screen for Child Anxiety Related Emotional Disorders; SD: standard deviation; STAI: State-Trait Anxiety Inventory

1 Downgraded 1 level - the cohort study by Lopez de Lara et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding and no control group).
2 Downgraded 1 level - the cohort study by Achille et al (2020) was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
3 Serious indirectness in Achille 2020- Outcome reported for full study cohort, of whom 30% were taking no treatment or puberty suppression alone at follow-up. Results for people taking gender-affirming hormones not reported separately.[4] Downgraded 1 level - the cohort study by Kuper et al. (2020) was assessed at high risk of bias (poor quality).
5 Downgraded 1 level - the cohort study by Allen et al. (2019) was assessed at high risk of bias (poor quality; lack of blinding and no control group).
6 Serious indirectness in Kuper et al. 2020- Outcome reported for full study cohort, of whom approximately 17% received puberty suppression alone and did not receive gender-affirming hormones
7 Downgraded 1 level - the cohort study by Kaltiala et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding and no control group).

Table 4: Question 1: For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Quality of life

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on quality of life (1 uncontrolled, prospective observational study and 1 uncontrolled, retrospective observational study)									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Change from baseline in mean quality of life score, measured using the QLES-Q-SF) (approximately 12-month follow-up). Higher scores indicated better quality of life.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=50	None	Numerical improvements in mean score reported from wave 1 (baseline) to wave 3 (approx. 12 months), but difference not statistically significant (p = 0.085) Results presented diagrammatically, numerical results for mean score not reported	Critical	VERY LOW
Change from baseline in adjusted mean well-being score, measured using the GWBS of the Pediatric Quality of Life Inventory (mean treatment duration 349 days). Higher scores indicated better well-being.									
1 cohort study Allen et al. 2019	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=39	None	T0 (baseline) = 61.70 (SE 2.43) T1 (final assessment) = 70.23 (SE 2.15) Statistically significant improvement in well-being score, p<0.002	Critical	VERY LOW

Abbreviations: GWBS: General Well-Being Scale; p: p-value; QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire; SE: standard error

1 Downgraded 1 level - the cohort study by Achille et al (2020) was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
2 Serious indirectness in Achille et al. 2020 - Outcome reported for full study cohort, of whom 30% were taking no treatment or puberty suppression alone at follow-up. Results for people taking gender-affirming hormones not reported separately.
3 Downgraded 1 level - the cohort study by Allen et al. (2019) was assessed at high risk of bias (poor quality; lack of blinding and no control group).

Table 5: Question 1: For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Body image

QUALITY	Summary of findings	IMPORTANCE	CERTAINTY

					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on body image *(1 uncontrolled, prospective observational study)*									
Change from baseline in mean body image, measured using the BIS (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores represent a higher degree of body dissatisfaction.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=86	None	Baseline = 70.7 (SD 15.2) Follow-up = 51.4 (SD 18.3) No statistical analysis reported for the sub-group of participants who received gender-affirming hormones	Important	VERY LOW

Abbreviations: BIS: Body Image Scale; p: p-value; SD: standard deviation

1 Downgraded 1 level - the cohort study by Kuper et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).

Table 6: Question 1: For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Psychological impact

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Psychosocial Impact (1 uncontrolled, prospective observational study and 1 uncontrolled, retrospective observational study)									
Change from baseline in family functioning, measured using the Family APGAR test. Higher scores suggest more family dysfunction.									
1 cohort study Lopez de Lara et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 17.9 T1 (12 months) = 18.0 No statistical analysis reported	Important	VERY LOW
Change from baseline in mean patient strengths and difficulties score, measured using the SDQ, Spanish Version (total difficulties score) (duration of treatment 12 months). Higher scores suggest the presence of a behavioural disorder.									
1 cohort study	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=23	None	T0 (baseline) = 14.7 (SD 3.3) T1 (12 months) = 10.3 (SD 2.9)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Lopez de Lara et al. 2020							Statistically significant improvement p<0.001		
Functioning in adolescent development: Living with parent(s)/ guardians[2] (outcome reported for the approximately 12-month period after starting gender-affirming hormones; referred to as the 'real-life phase' in Finland). Not living with parent(s) or guardian in your early 20s is a marker of age-appropriate functioning in Finnish culture.									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During gender identity assessment = 73% (38/52) During real life phase = 40% (21/50) Statistically significant reduction (p=0.001)	Important	VERY LOW
Functioning in adolescent development: Normative peer contacts[4] (outcome reported for the approximately 12-month period after starting gender-affirming hormones; referred to as the 'real-life phase' in Finland)									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During gender identity assessment = 89% (46/52) During real life phase = 81% (42/52) Statistically significant reduction (p<0.001)	Important	VERY LOW
Functioning in adolescent development: Progresses normatively in school/ work[5] (outcome reported for the approximately 12-month period after starting gender-affirming hormones; referred to as the 'real-life phase' in Finland)									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During gender identity assessment = 64% (33/52) During real life phase = 60% (31/52) No statistically significant difference (p=0.69)	Important	VERY LOW
Functioning in adolescent development: Has been dating or had steady relationships[6] (outcome reported for the approximately 12-month period after starting gender-affirming hormones; referred to as the 'real-life phase' in Finland)									
1 cohort study	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During gender identity assessment = 62% (32/50)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Kaltiala et al. 2020							During real life phase = 58% (30/52) No statistically significant difference (p=0.51)		
Functioning in adolescent development: Is age-appropriately able to deal with matters outside of the home[7] (outcome reported for the approximately 12-month period after starting gender-affirming hormones; referred to as the 'real-life phase' in Finland)									
1 cohort study Kaltiala et al. 2020	Serious limitations[2]	No serious indirectness	No serious inconsistency	Not calculable	N=52	None	During gender identity assessment = 81% (42/52) During real life phase = 81% (42/52) No statistically significant difference (p=1.00)	Important	VERY LOW

Abbreviations: APGAR: Adaptability, Partnership, Growth, Affection and Resolve; p: p-value; SD: standard deviation; SDQ: Strengths and Difficulties Questionnaire

1 Downgraded 1 level - the cohort study by Lopez de Lara et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding and no control group).
2 Living arrangements were classified as (1) living with at least one parent/guardian, (2) living in a boarding school, with an adult relative, in some form of supported accommodation or the like, where supervision and guidance by a responsible adult is provided, (3) independently alone or in a shared household with a peer, (4) with a romantic partner. In the analyses dichotomised living arrangements as (a) parent(s)/guardian(s) vs. in other arrangements.
3 Downgraded 1 level - the cohort study by Kaltiala et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding and no control group).
4 Peer relationships were classified as: (1) socialises with friends in leisure time, outside of activities supervised by adults, (2) socialises with peers only at school or in the context of rehabilitative activity, (3) spends time close to peers, for example in school or rehabilitative activity, but does not connect with them, (4) does not meet peers at all. In the analyses, peer relationships during (a) gender identity assessment and (b) the real-life phase were dichotomized to age-appropriate (normative) (1) vs. restricted or lacking (2–4).
5 School/work participation was classified as (1) age appropriate participation in mainstream curriculum, progresses without difficulties, (2) participates in mainstream curriculum with difficulty, (3) participates in rehabilitative educational or work activity, (4) not involved in education and working life. Age-appropriate participation during (1) was recorded if the adolescent attended mainstream secondary education or upper secondary education at a regular rate (a class per year in comprehensive school; has not changed more than once between tracks in upper secondary education) or had proceeded to work life after completing vocational education. Participation with difficulty (2) was recorded if the adolescent was enrolled in mainstream education but had to repeat a class, studied with special arrangements (for example, in a special small group), or followed some form of adjusted curriculum. In the analyses, school/work life during (a) gender identity assessment and (b) real-life phase was dichotomised to normative (1) vs. any other (2, 3 or 4).
6 Romantic involvement was recorded (1) has or has had a dating or steady relationship, not only online, (2) has had a romantic relationship only online, (3) has not had dating or steady relationships. In the analyses we compared has or has had (1) vs. has not had (2,3) a dating or steady relationship during (a) gender identity assessment and (b) real-life phase. Sexual history was recorded in more detail in case histories during gender identity assessment, and for this period we also collected the experiences of (French) kissing (yes/no), intercourse (yes/no) and experience of any genitally intimate contact with a partner (petting under clothes or naked, intercourse, oral sex) (yes/no).

7 In recording age-appropriate competence in managing everyday matters it was expected that early adolescents (up to 14 years) would be able, for example, to do shopping and travel alone on local public transport, and to help with household duties assigned by their parents. Middle adolescents (15–17 years) were further assumed, for example, to be able make telephone calls in matters important to them (for example, when seeking a summer job), to deal with school-related issues with school personnel without parental participation, to select and start new hobbies independently and to fulfil their role in summer jobs and in similar responsibilities of young people. Late adolescents (18 years and over), legally adults, were expected to have, in addition to the above, competence to talk to authorities such as professionals in health and social services, employment or educational institutions, to deal with banks or health insurance, to manage their financial issues and to manage their housekeeping if they chose to move to live independently of parents/guardians. Competence in managing everyday matters was recorded as follows: (1) the adolescent is able to cope age appropriately outside home, (2) the adolescent needs support in age-appropriate matters outside home but functions age-appropriately in the home (manages her/his own hygiene, clothing and nutrition, participates in (younger subjects) or takes responsibility for (older subjects) housekeeping) and (3) the adolescent's functioning is inadequate both at home and outside home. For the analyses, participants were determined to be able to age-appropriately able cope with matters outside of the home (1) vs. not (2,3).

Table 7: Question 2: For children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Bone density

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Lumbar spine bone mineral apparent density (BMAD) (2 uncontrolled, retrospective observational studies)									
Change from start of gender-affirming hormones to age 22 years in lumber spine BMAD in transfemales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=13 (Mean) N=14 (z-score)	None	Mean (SD), g/m^3 Start of gender-affirming hormones: 0.22 (0.02) Age 22 years: 0.23 (0.03) P=0.003 z-score (SD) Start of gender-affirming hormones: -0.90 (0.80) Age 22 years: -0.78 (1.03) No statistically significant difference	Important	VERY LOW
Change from baseline in lumbar spine BMAD in transfemales with a bone age less than 15 years ('young'; 24 months follow-up)									
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=15	None	Median (range), g/m^3 Start of gender-affirming hormones (C0): 0.20 (0.18 to 0.24)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							24-month follow-up (C24): 0.22 (0.19 to 0.27) Statistically significant increase (p≤0.01) z-score (range) Start of gender-affirming hormones (C0): -1.52 (-2.36 to 0.42) 24-month follow-up (C24): -1.10 (-2.44 to 0.69) Statistically significant increase (p≤0.05)		
Change from baseline in lumbar spine BMAD in transfemales with a bone age of 15 years or more ('old'; 24 months follow-up)									
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=5	None	Median (range), g/m^3 Start of gender-affirming hormones (C0): 0.22 (0.19 to 0.24) 24-month follow-up (C24): 0.23 (0.21 to 0.26) Statistically significant increase (p≤0.05) z-score (range) Start of gender-affirming hormones (C0): -1.15 (-2.21 to 0.08) 24-month follow-up (C24): -0.66 (-1.66 to 0.54) Statistically significant increase (p≤0.05)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in lumber spine BMAD in transmales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=19 (Mean and z-score)	None	Mean (SD), g/m^3 Start of gender-affirming hormones: 0.24 (0.02) Age 22 years: 0.25 (0.28) P=0.001 z-score Start of gender-affirming hormones: -0.50 (0.81) Age 22 years: -0.033 (0.95) P=0.002	Important	VERY LOW
Change from baseline in lumbar spine BMAD in transmales with a bone age of less than 14 years ('young'; 24 months follow-up)									
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=11	None	Median (range), g/m^3 Start of gender-affirming hormones (C0): 0.23 (0.19 to 0.28) 24-month follow-up (C24): 0.25 (0.22 to 0.28) Statistically significant increase (p≤0.01) z-score (range) Start of gender-affirming hormones (C0): -0.84 (-2.2 to 0.87) 24-month follow-up (C24): -0.15 (-1.38 to 0.94) Statistically significant increase (p≤0.01)	Important	VERY LOW
Change from baseline in lumbar spine BMAD in transmales with a bone age of 14 years or more ('old'; 24 months follow-up)									
1 cohort study	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=23	None	Median (range), g/m^3	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Vlot et al. 2017							Start of gender-affirming hormones (C0): 0.24 (0.20 to 0.28) 24-month follow-up (C24): 0.25 (0.21 to 0.30) Statistically significant increase (p≤0.01) z-score (range) Start of gender-affirming hormones (C0): -0.29 (-2.28 to 0.90) 24-month follow-up (C24): -0.06 (-1.75 to 1.61) Statistically significant increase (p≤0.01)		
Change in femoral neck BMAD (2 uncontrolled, retrospective observational studies)									
Change from start of gender-affirming hormones to age 22 years in femoral neck BMAD in transfemales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=14 (Mean) N=10 (z-score)	None	Mean (SD), g/m^3 Start of gender-affirming hormones: 0.26 (0.04) Age 22 years: 0.28 (0.05) No statistically significant difference z-score (SD) Start of gender-affirming hormones: -1.57 (1.74) Age 22 years: Not reported	Important	VERY LOW
Change from baseline in femoral neck BMAD in transfemales with a bone age less than 15 years ('young'; 24 months follow-up)									
1 cohort study	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=16	None	Median (range), g/m^3 C0: 0.27 (0.20 to 0.33) C24: 0.27 (0.20 to 0.36)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Vlot et al. 2017							No statistically significant change z-score (range) C0: -1.32 (-3.39 to 0.21) C24: -1.30 (-3.51 to 0.92) No statistically significant change		
Change from baseline in femoral neck BMAD in transfemales with a bone age of 15 years or more ('old'; 24 months follow-up)									
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=6	None	Median (range), g/m^3 C0: 0.30 (0.26 to 0.34) C24: 0.29 (0.24 to 0.38) No statistically significant change z-score (range) C0: -0.36 (-1.50 to 0.46) C24: -0.56 (-2.17 to 1.29) No statistically significant change	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in femoral neck BMAD in transmales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=19 (Mean) N=18 (z-score)	None	Mean (SD), g/m^3 Start of gender-affirming hormones: 0.31 (0.04) Age 22 years: 0.33 (0.05) P=0.010 z-score (SD) Start of gender-affirming hormones: -0.28 (0.74) Age 22 years: Not reported	Important	VERY LOW
Change from baseline in femoral neck BMAD in transmales with a bone age of less than 14 years ('young'; 24 months follow-up)									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=10	None	Median (range), g/m^3 C0: 0.30 (0.22 to 0.35) C24: 0.33 (0.23 to 0.37) Statistically significant increase (p≤0.01) z-score (range) C0: -0.37 (-2.28 to 0.47) C24: -0.37 (-2.03 to 0.85) Statistically significant increase (p≤0.01)	Important	VERY LOW
Change from baseline in femoral neck BMAD in transmales with a bone age of 14 years or more ('old'; 24 months follow-up)									
1 cohort study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=23	None	Median (range), g/m^3 C0: 0.30 (0.23 to 0.41) C24: 0.32 (0.23 to 0.41) Statistically significant increase (p≤0.01) z-score (range) C0: -0.27 ((-1.91 to 1.29) C24: 0.02 (-2.1 to 1.35) Statistically significant increase (p≤0.05)	Important	VERY LOW
Change in lumbar spine BMD (2 uncontrolled, retrospective observational studies)									
Change from start of gender-affirming hormones to age 22 years in lumbar spine BMD in transfemales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=15 (Mean) N=13 (z-score)	None	Mean (SD), g/m^2 Start of gender-affirming hormones: 0.84 (0.11) Age 22 years: 0.93 (0.10) P<0.001 z-score (SD)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							Start of gender-affirming hormones: -1.01 (0.98) Age 22 years: -1.36 (0.83) No statistically significant difference		
Change from start of gender-affirming hormones to age 22 years in lumbar spine BMD in transmales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=19 (Mean and z-score)	None	Mean (SD), g/m^2 Start of gender-affirming hormones: 0.91 (0.10) Age 22 years: 0.99 (0.13) P<0.001 z-score (SD) Start of gender-affirming hormones: -0.72 (0.99) Age 22 years: -0.33 (1.12) No statistically significant difference	Important	VERY LOW
Change from start of testosterone treatment in lumbar spine BMD in transmen (follow-up 6 to 24 months)									
1 cohort study Stoffers et al. 2019	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable	N=62 (T0 and T6) N=37 (T12) N=15 (T24)	None	Mean (SD), g/cm^2 T0: 0.90 (0.11) T6: 0.94 (0.10) T12: 0.95 (0.09) T24: 0.95 (0.11) No statistically significant difference from T0 to any timepoint z-score (SD) T0: -0.81 (1.02) T6: -0.67 (0.95) T12: -0.66 (0.81) T24: -0.74 (1.17)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							No statistically significant difference from T0 to any timepoint		
Change in femoral neck BMD (2 uncontrolled, retrospective observational studies)									
Change from start of gender-affirming hormones to age 22 years in femoral neck BMD in transfemales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=15 (Mean) N=11 (z-score)	None	Mean (SD), g/m^2 Start of gender-affirming hormones: 0.87 (0.08) Age 22 years: 0.94 (0.11) P=0.009 z-score (SD) Start of gender-affirming hormones: -0.95 (0.63) Age 22 years: -0.69 (0.74) No statistically significant difference	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in femoral neck BMD in transmales									
1 cohort study Klink et al. 2015	Serious limitations[1]	Serious indirectness[2]	Not applicable	Not calculable	N=19 (Mean) N=16 (z-score)	None	Mean (SD), g/m^2 Start of gender-affirming hormones: 0.88 (0.09) Age 22 years: 0.95 (0.10) P<0.001 z-score (SD) Start of gender-affirming hormones: -0.35 (0.79) Age 22 years: -0.35 (0.74) P=0.006	Important	VERY LOW
Change from start of testosterone treatment in right femoral neck (hip) BMD in transmales (follow-up 6 to 24 months)									
1 cohort study	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable	N=62 (T0 and T6)	None	Mean (SD), g/cm^2 T0: 0.77 (0.08)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Stoffers et al. 2019					N=37 (T12) N=15 (T24)		T6: 0.84 (0.11) T12: 0.82 (0.08) T24: 0.85 (0.11) No statistically significant difference from T0 to any timepoint z-score (SD) T0: -0.97 (0.79) T6: -0.54 (0.96) T12: -0.80 (0.69) T24: -0.31 (0.84) No statistically significant difference from T0 to any timepoint		
Change from start of testosterone treatment in left femoral neck (hip) BMD in transmales (follow-up 6 to 24 months)									
1 cohort study Stoffers et al. 2019	Serious limitations[4]	No serious indirectness	Not applicable	Not calculable	N=62 (T0 and T6) N=37 (T12) N=15 (T24)	None	Mean (SD), g/cm^2 T0: 0.76 (0.09) T6: 0.83 (0.12) T12: 0.81 (0.08) T24: 0.86 (0.09) No statistically significant difference from T0 to any timepoint z-score (SD) T0: -1.07 (0.85) T6: -0.62 (1.12) T12: -0.93 (0.63) T24: -0.20 (0.70) No statistically significant difference from T0 to any timepoint	Important	VERY LOW

Abbreviations: BMAD: bone mineral apparent density; BMD: bone mineral density; g: grams; m: metre; SD: standard deviation

1 Downgraded 1 level - the cohort study by Klink et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding, no control group and high number of participants lost to follow-up)
2 Outcomes reported after gender reassignment surgery and not after gender-affirming hormones alone. Unclear whether observed changes are due to hormones or surgery
3 Downgraded 1 level - the cohort study by Vlot et al. (2017) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control)
4 Downgraded 1 level - the cohort study by Stoffers et al. (2019) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group)

Table 8: Question 2: For children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Cardiovascular risk factors

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change in body mass index (1 uncontrolled, retrospective observational study)									
Change from start of gender-affirming hormones to age 22 years in BMI in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI) +1.9 (0.6 to 3.2) Statistically significant increase (p<0.005) Mean BMI at 22 years (95% CI): 23.2 (21.6 to 24.8)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in BMI in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI) +1.4 (0.8 to 2.0) Statistically significant increase (p<0.005) Mean BMI at 22 years (95% CI): 23.9 (23.0 to 24.7)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Obesity rates at age 22 years (1 uncontrolled, retrospective observational study)									
Obesity rates at age 22 years in transfemales who started gender-affirming hormones as adolescents (1 uncontrolled, retrospective observational study)									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	At 22 years, 9.9% of transfemales were obese, compared with 3.0% in reference cisgender population No statistically analysis reported	Important	VERY LOW
Obesity rates at age 22 years in transfemales who started gender-affirming hormones as adolescents (1 uncontrolled, retrospective observational study)									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	At 22 years, 6.6% of transmales were obese, compared with 2.2% in reference cisgender population No statistically analysis reported	Important	VERY LOW
Change in blood pressure (1 uncontrolled, retrospective observational study)									
Change from start of gender-affirming hormones to age 22 years in systolic blood pressure (SBP) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI) -3 (-8 to 2) No statistically significant difference Mean SBP at 22 years (95% CI): 117 (113 to 122)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in diastolic blood pressure (DBP) in transfemales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI) +6 (3 to 10) Statistically significant increase (p<0.001) Mean DBP at 22 years (95% CI): 75 (72 to 78)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in systolic blood pressure (SBP) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI): +5 (1 to 9) Statistically significant increase (p<0.05) Mean SBP at 22 years (95% CI): 126 (122 to 130)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in diastolic blood pressure (DBP) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI): +6 (4 to 9) Statistically significant increase (p<0.001) Mean DBP at 22 years (95% CI): 74 (72 to 77)	Important	VERY LOW
Change in glucose levels, insulin levels, insulin resistance and HbA1c (2 uncontrolled, retrospective observational studies)									
Change from start of gender-affirming hormones to age 22 years in glucose level (mmol/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI): +0.1 (-0.1 to 0.2)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							No statistically significant difference Mean glucose level at 22 years (95% CI): 5.0 (4.8 to 5.1)		
Change from start of gender-affirming hormones to age 22 years in insulin level (mU/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI) +2.7 (-1.7 to 7.1) No statistically significant difference Mean insulin level at 22 years (95% CI): 13.0 (8.4 to 17.6)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in insulin resistance (HOMA-IR) in transfemales. Higher scores indicate more insulin resistance.									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI) +0.7 (-0.2 to 1.5) No statistically significant difference Mean HOMA-IR at 22 years (95% CI): 2.9 (1.9 to 3.9)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in glucose level (mmol/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI) 0.0 (-0.2 to 0.2) No statistically significant difference	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							Mean glucose level at 22 years (95% CI): 4.8 (4.7 to 5.0)		
Change from start of gender-affirming hormones to age 22 years in insulin level (mU/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI) -2.1 (-3.9 to -0.3) Statistically significant decrease (p<0.05) Mean insulin level at 22 years (95% CI): 8.6 (6.9 to 10.2)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in insulin resistance (HOMA-IR) in transmales. Higher scores indicate more insulin resistance.									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI): -0.5 (-1.0 to -0.1) Statistically significant decrease (p<0.05) Mean HOMA-IR at 22 years (95% CI): 1.8 (1.4 to 2.2)	Important	VERY LOW
Change from start of testosterone in HbA1c in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N= Not reported	None	No statistically significant change from start of testosterone treatment Numerical results, follow-up duration and further details of	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							statistical analysis not reported.		
Change in lipid profile (1 uncontrolled, retrospective observational study)									
Change from start of gender-affirming hormones to age 22 years in total cholesterol (mmol/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI): +0.1 (-0.2 to 0.4) No statistically significant difference Mean total cholesterol at 22 years (95% CI): 4.1 (3.8 to 4.4)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in HDL cholesterol (mmol/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI): 0.0 (-0.1 to 0.2) No statistically significant difference Mean HDL cholesterol at 22 years (95% CI): 1.6 (1.4 to 1.7)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in LDL cholesterol (mmol/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI): 0.0 (-0.3 to 0.2) No statistically significant difference Mean LDL cholesterol at 22 years (95% CI): 2.0 (1.8 to 2.3)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change from start of gender-affirming hormones to age 22 years in triglycerides (mmol/L) in transfemales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=71	None	Mean change (95% CI): +0.2 (0.0 to 0.5) Statistically significant increase (p<0.05) Mean triglycerides at 22 years (95% CI): 1.1 (0.9 to 1.4)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in total cholesterol (mmol/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI): +0.4 (0.2 to 0.6) Statistically significant increase (p<0.001) Mean total cholesterol at 22 years (95% CI): 4.6 (4.3 to 4.8)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in HDL cholesterol (mmol/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI) -0.3 (-0.4 to -0.2) Statistically significant decrease (p<0.001) Mean HDL cholesterol at 22 years (95% CI): 1.3 (1.2 to 1.3)	Important	VERY LOW
Change from start of gender-affirming hormones to age 22 years in LDL cholesterol (mmol/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI): +0.4 (0.2 to 0.6)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
							Statistically significant increase (p<0.001) Mean LDL cholesterol at 22 years (95% CI): 2.6 (2.4 to 2.8)		
Change from start of gender-affirming hormones to age 22 years in triglycerides (mmol/L) in transmales									
1 cohort study Klaver et al. 2020	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=121	None	Mean change (95% CI) +0.5 (0.3 to 0.7) Statistically significant increase (p<0.001) Mean triglycerides at 22 years (95% CI): 1.3 (1.1 to 1.5)	Important	VERY LOW

Abbreviations: BMI: boss mass index; CI: confidence interval; DBP: diastolic blood pressure; HbA1c: glycated haemoglobin; HDL: high-density lipoproteins; HOMA-IR: Homeostatic Model Assessment of Insulin Resistance; LDL: low-density lipoproteins; mmol/L: millimoles per litre; mU/L: milliunits per litre; SBP: systolic blood pressure; SD: standard deviation

1 Downgraded 1 level - the cohort study by Klaver et al. (2020) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group)
2 Downgraded 1 level - the cohort study by Stoffers et al. (2019) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group)

Table 9: Question 2: For children and adolescents with gender dysphoria, what is the short-term and long-term safety of gender-affirming hormones compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – Other safety outcomes

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Liver enzymes (1 uncontrolled, retrospective observational study)									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change from start of testosterone in aspartate aminotransferase (AST) level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N= Not reported	None	No statistically significant change from start of testosterone treatment Numerical results, follow-up duration and further details of statistical analysis not reported.	Important	VERY LOW
Change from start of testosterone in alanine aminotransferase (ALT) level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N= Not reported	None	No statistically significant change from start of testosterone treatment Numerical results, follow-up duration and further details of statistical analysis not reported.	Important	VERY LOW
Change from start of testosterone in gamma-glutamyl transferase (GGT) level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N= Not reported	None	No statistically significant change from start of testosterone treatment Numerical results, follow-up duration and further details of statistical analysis not reported.	Important	VERY LOW
Change from start of testosterone in alkaline phosphatase (ALP) level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=62 (T0 and T1) N=37 (T12)	None	Median (IQR), U/L T0: 102 (78 to 136) T6: 115 (102 to 147) T12: 112 (88 to 143) T24: 81 (range 69 to 98)	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
					N-15 (T24)		Statistically significant increase from T0 at T6 and T12 (p<0.001)		
Kidney markers (1 uncontrolled, retrospective observational study)									
Change from start of testosterone in serum creatinine level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=62 (T0 and T1) N=37 (T12) N=15 (T24)	None	Mean (SD), umol/L T0: 62 (7) T6: 70 (9) T12: 74 (10) T24: 81 (10) Statistically significant increase from T0 at all timepoints (p<0.001)	Important	VERY LOW
Change from start of testosterone in serum urea[2] level in transmales (up to 24 months follow-up)									
1 cohort study Stoffers et al. 2019	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N= Not reported	None	No statistically significant change from start of testosterone treatment Numerical results, follow-up duration and further details of statistical analysis not reported.	Important	VERY LOW
Adverse effects (1 uncontrolled, retrospective observational study)									
Permanent discontinuation of gender-affirming hormones (median follow-up 2.0 years (range 0.0 to 11.3)									
1 cohort study Khatchadourian et al. 2014	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=63	None	No participants permanently discontinued gender-affirming hormones.	Important	VERY LOW
Temporary discontinuation of gender-affirming hormones (median follow-up 2.0 years (range 0.0 to 11.3)									
1 cohort study	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=63	None	3/37 transmales receiving testosterone temporarily	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Khatchadourian et al. 2014							discontinued treatment, 2 due to concomitant mental health comorbidities and 1 due to androgenic alopecia. All eventually resumed treatment. No transfemales receiving oestrogen temporarily discontinued treatment		
Minor complications during treatment with gender-affirming hormones (median follow-up 2.0 years (range 0.0 to 11.3)									
1 cohort study Khatchadourian et al. 2014	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=63	None	12/63 participants had minor complications during treatment with gender-affirming hormones All 12 were transmales receiving testosterone. Complications were severe acne (n=7), androgenic alopecia (n=1) mild dyslipidaemia (n=3) and significant mood swings (n=1) No transfemales receiving oestrogen had minor complications	Important	VERY LOW
Severe complications during treatment with gender-affirming hormones (median follow-up 2.0 years (range 0.0 to 11.3)									
1 cohort study Khatchadourian et al. 2014	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=63	None	No severe complications reported during gender-affirming treatment	Important	VERY LOW

Abbreviations: ALP: alkaline phosphatase; ALT: alanine aminotransferase; AST: aspartate aminotransferase; GGT: gamma-glutamyl transferase; IQR: interquartile range; SD: standard deviation; U/L: units per litre; umol/L: micromole per litre

1 Downgraded 1 level - the cohort study by Stoffers et al. (2019) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group)
2 Referred to as 'ureum' in original publication
3 Downgraded 1 level - the cohort study by Khatchadourian et al. (2014) was assessed as at high risk of bias (poor quality overall; lack of blinding, no control group and high number of participants lost to follow-up)

Table 10: From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria? – Transfemales compared with transmales

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Transfemales	Transmales	Result (95% CI)		
Impact on mental health (1 uncontrolled, retrospective observational study)									
Change from baseline in adjusted mean suicidality score, measured using the ASQ tool (mean treatment duration 349 days). Higher scores indicate a greater degree of suicidality.									
1 cohort study Allen et al. 2019	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=14	N=33	**Transfemales** T0 (baseline) = 1.21 (SE 0.36) T1 (final assessment) = 0.24 (SE 0.19) **Transmales** T0 (baseline) = 1.01 (SE 0.23) T1 (final assessment) = 0.29 (SE 0.13) No statistically significant difference in change from baseline between transfemales and transmales (p=0.79)	Critical	VERY LOW
Impact on quality of life (1 uncontrolled, retrospective observational study)									
Change from baseline in adjusted mean well-being score, measured using the GWBS of the Pediatric Quality of Life Inventory (mean treatment duration 349 days). Higher scores indicate better well-being.									
1 cohort study Allen et al. 2019	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	N=14	N=33	**Transfemales** T0 (baseline) = 58.44 (SE 4.09) T1 (final assessment) = 69.52 (SE 3.62)	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Transfemales	Transmales	Result (95% CI)		
							Transmales T0 (baseline) = 64.95 (SE 2.66) T1 (final assessment) = 70.94 (SE 2.35) No statistically significant difference in change from baseline between transfemales and transmales (p=0.32)		

Abbreviations: ASQ: Ask Suicide-Screening Questions; GWBS: General Well-Being Scale; SE: standard error

1 The cohort study by Allen et al. 2019 was assessed at high risk of bias (poor quality; lack of blinding and no control group).

Table 11: From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria? – Sex assigned at birth males (transfemales)

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study type and number of studies Author year	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change from baseline in mean depression symptoms in transfemales, measured using the Quick Inventory of Depressive Symptoms (QIDS), self-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more depression.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=40	None	Baseline = 7.5 (SD 4.9) Follow-up = 6.6 (SD 4.4) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean depression symptoms in transfemales, measured using the Quick Inventory of Depressive Symptoms (QIDS), clinician-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe depression.									
1 cohort study	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=45	None	Baseline = 4.2 (SD 3.2) Follow-up = 5.4 (SD 3.4)	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study type and number of studies Author year	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Kuper et al. 2020							No statistical analysis reported for this sub-group		
Change from baseline in mean anxiety symptoms in transfemales, measured using the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe anxiety.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=33	None	Baseline = 26.4 (SD 14.2) Follow-up = 24.3 (SD 15.4) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean panic symptoms in transfemales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=34	None	Baseline = 5.7 (SD 4.9) Follow-up = 5.1 (SD 4.9) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean generalised anxiety symptoms in transfemales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=34	None	Baseline = 8.6 (SD 5.1) Follow-up = 8.0 (SD 5.1) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean social anxiety symptoms in transfemales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=34	None	Baseline = 7.1 (SD 3.9) Follow-up = 6.8 (SD 4.4) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean separation anxiety symptoms in transfemales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=34	None	Baseline = 3.4 (SD 3.3) Follow-up = 2.7 (SD 2.3) No statistical analysis reported for this sub-group	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study type and number of studies Author year	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change from baseline in mean school avoidance symptoms in transfemales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=33	None	Baseline = 1.8 (SD 1.7) Follow-up = 1.9 (SD 2.1) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in percentage of participants with suicidal ideation in transfemales, measured using the additional questions from the PHQ 9_Modified for Teens (approximately 12-month follow-up)									
1 cohort study Achille et al. 2020	Serious limitations[2]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=17	None	Wave 1 (baseline) = 11.8% (2/17) Wave 2 (approx. 12 months) = 5.9% (1/17) No statistical analysis reported	Critical	VERY LOW
Impact on body image (1 uncontrolled, prospective observational study)									
Change from baseline in mean body image in transfemales, measured using the BIS (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores represent a higher degree of body dissatisfaction.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=30	None	Baseline = 67.5 (SD 19.5) Follow-up = 49.0 (SD 21.6) No statistical analysis reported for this sub-group	Important	VERY LOW

Abbreviations: BIS: Body Image Scale; PHQ 9: Patient Health Questionnaire 9; SCARED: Screen for Child Anxiety Related Emotional Disorders; SD: standard deviation

1 Downgraded 1 level - the cohort study by Kuper et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
2 Downgraded 1 level - the cohort study by Achille et al. 2020 was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
3 Serious indirectness in Achille 2020- Approximately 30% of the full sample received puberty suppression alone or were receiving no treatment at final follow-up.

Table 12: From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria? – Sex assigned at birth females (transmales)

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change from baseline in mean depression symptoms in transmales, measured using the Quick Inventory of Depressive Symptoms (QIDS), self-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe depression.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=76	None	Baseline = 10.4 (SD 5.0) Follow-up = 7.5 (SD 4.5) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean depression symptoms in transmales, measured using the Quick Inventory of Depressive Symptoms (QIDS), clinician-reported (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe depression.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=78	None	Baseline = 6.7 (SD 4.4) Follow-up = 6.2 (SD 4.1) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean anxiety symptoms in transmales, measured using the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe anxiety.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=65	None	Baseline = 35.4 (SD 16.5) Follow-up = 29.8 (SD 15.5) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean panic symptoms in transmales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=66	None	Baseline = 9.3 (SD 6.5) Follow-up = 7.9 (SD 6.5) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean generalised anxiety symptoms in transmales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=66	None	Baseline = 10.4 (SD 5.0) Follow-up = 9.0 (SD 5.1) No statistical analysis reported for this sub-group	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change from baseline in mean social anxiety symptoms in transmales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=66	None	Baseline = 8.5 (SD 4.0) Follow-up = 7.8 (SD 4.1) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean separation anxiety symptoms in transmales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=65	None	Baseline = 4.2 (SD 3.4) Follow-up = 3.4 (SD 2.6) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in mean school avoidance symptoms in transmales, measured using specific questions from the SCARED questionnaire (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores indicate more severe symptoms.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=65	None	Baseline = 2.9 (SD 2.3) Follow-up = 2.0 (SD 2.3) No statistical analysis reported for this sub-group	Critical	VERY LOW
Change from baseline in percentage of participants with suicidal ideation in transmales, measured using the additional questions from the PHQ 9_Modified for Teens (approximately 12-month follow-up)									
1 cohort study Achille et al. 2020	Serious limitations[2]	Serious indirectness[3]	No serious inconsistency	Not calculable	N=33	None	Wave 1 (baseline) = 9.1% (3/33) Wave 2 (approx. 12 months) = 6.1% (2/33) No statistical analysis reported	Critical	VERY LOW
Impact on body image (1 uncontrolled, prospective observational study)									
Change from baseline in mean body image in transmales, measured using the BIS (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores represent a higher degree of body dissatisfaction.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=66	None	Baseline = 71.1 (SD 13.4) Follow-up = 52.9 (SD 16.8) No statistical analysis reported for this sub-group	Important	VERY LOW

Abbreviations: BIS: Body Image Scale; PHQ 9: Patient Health Questionnaire 9; SCARED: Screen for Child Anxiety Related Emotional Disorders; SD: standard deviation

1 Downgraded 1 level - the cohort study by Kuper et al. (2020) was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
2 Downgraded 1 level - the cohort study by Achille et al. 2020 was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
3 Serious indirectness in Achille 2020- Approximately 30% of the full sample received puberty suppression alone or were receiving no treatment at final follow-up.

Table 14: From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria? – Outcomes controlled for concurrent counselling and medicines for mental health problems

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Impact on mental health (1 uncontrolled, retrospective observational study)									
Change from baseline in mean depression score in transfemales, measured using the CESD-R (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicate more depression.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=17	None	No statistically significant change from baseline (p=0.27) Numerical scores not reported	Critical	VERY LOW
Change from baseline in mean depression score in transmales, measured using the CESD-R (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicate more severe depression.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=33	None	No statistically significant change from baseline (p=0.43) Numerical scores not reported	Critical	VERY LOW
Change from baseline in depression score in transfemales, measured using the Patient Health Questionnaire Modified for Teens (PHQ 9_Modified for Teens) (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicate more severe depression.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=17	None	No statistically significant change from baseline (p=0.07) Numerical scores not reported	Critical	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Change from baseline in depression score in transmales, measured using the Patient Health Questionnaire Modified for Teens (PHQ 9_Modified for Teens) (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicate more severe depression.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=33	None	No statistically significant change from baseline (p=0.67) Numerical scores not reported	Critical	VERY LOW
Impact on quality of life (1 uncontrolled, retrospective observational study)									
Change from baseline in mean quality of life score in transfemales, measured using the QLES-Q-SF (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicated better quality of life.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=17	None	No statistically significant change from baseline (p=0.06)	Critical	VERY LOW
Change from baseline in mean quality of life score in transmales, measured using the QLES-Q-SF (approximately 12-month follow-up; controlled for engagement in counselling and medicines for mental health problems). Higher scores indicated better quality of life.									
1 cohort study Achille et al. 2020	Serious limitations[1]	Serious indirectness[2]	No serious inconsistency	Not calculable	N=33	None	No statistically significant change from baseline (p=0.08)	Critical	VERY LOW
Psychosocial Impact (1 uncontrolled, retrospective observational study)									
Functioning in adolescent development: Progresses normatively in school/ work during the real-life phase – impact on need for mental health treatment before or during gender identity assessment									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=49	None	Needed mental health treatment: 47% (15/32) functioning well Did not need mental health treatment: 82% (14/17) functioning well Statistically significant difference p=0.02	Important	VERY LOW
Functioning in adolescent development: Is age-appropriately able to deal with matters outside of the home during the real-life phase – impact on need for mental health treatment before or during gender identity assessment									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=49	None	Needed mental health treatment: 72% (23/32) managing well Did not need mental health treatment: 94% (16/17) managing well No statistically significant difference p=0.06	Important	VERY LOW
Functioning in adolescent development: Progresses normatively in school/ work during the real-life phase – impact on need for mental health treatment during the real-life phase									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=51	None	Needed mental health treatment: 42% (10/24) functioning well Did not need mental health treatment: 74% (20/27) functioning well Statistically significant difference p=0.02	Important	VERY LOW
Functioning in adolescent development: Is age-appropriately able to deal with matters outside of the home during the real-life phase – impact on need for mental health treatment during the real-life phase									
1 cohort study Kaltiala et al. 2020	Serious limitations[3]	No serious indirectness	No serious inconsistency	Not calculable	N=51	None	Needed mental health treatment: 67% (16/24) managing well Did not need mental health treatment: 93% (25/27) managing well Statistically significant difference p=0.02	Important	VERY LOW

Abbreviations: CESD-R: Center for Epidemiologic Studies Depression; p: p-value; PHQ 9: Patient Health Questionnaire 9; QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire

1 Downgraded 1 level - the cohort study by Achille et al 2020 was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).
2 Serious indirectness in Achille 2020- Approximately 30% of the full sample received puberty suppression alone or were receiving no treatment at final follow-up.
3 Downgraded 1 level - the cohort study by Kaltiala et al. 2020 was assessed at high risk of bias (poor quality; lack of blinding and no control).

Table 15: From the evidence selected, are there particular sub-groups of children and adolescents with gender dysphoria that derive comparatively more (or less) benefit from treatment with gender-affirming hormones than the wider population of children and adolescents with gender dysphoria? – Tanner age

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of patients		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result (95% CI)		
Impact on mental health (1 uncontrolled, retrospective observational study)									
Change from baseline in mental health problems – depression, anxiety and anxiety-related symptoms (mean duration of gender-affirming hormone treatment was 10.9 months)									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=105	None	No difference in outcomes found by Tanner age. Numerical results, statistical analysis and information on specific outcomes not reported. It is unclear from the paper whether Tanner age is at initial assessment, start of GnRH analogues, start of gender-affirming hormones, or another timepoint	Critical	VERY LOW
Impact on body image (1 uncontrolled, prospective observational study)									
Change from baseline in mean body image, measured using the BIS (mean duration of gender-affirming hormone treatment was 10.9 months). Higher scores represent a higher degree of body dissatisfaction.									
1 cohort study Kuper et al. 2020	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=105	None	No difference in body image score found by Tanner age. Numerical results, statistical analysis and information on specific outcomes not reported.	Important	VERY LOW

							It is unclear from the paper whether Tanner age is at initial assessment, start of GnRH analogues, start of gender-affirming hormones, or another timepoint		

Abbreviations: BIS: Body Image Scale

1 Downgraded 1 level - the cohort study by Kuper et al. 2020 was assessed at high risk of bias (poor quality; lack of blinding, no control group and high number of participants lost to follow-up).

Glossary

Ask Suicide-Screening Questions (ASQ)	ASQ is a four-item dichotomous (yes, no) response measure with high sensitivity, designed to identify risk of suicide. A patient is considered to have screened positive if they answered yes to any item. The authors of Allen et al. 2019 altered the fourth item of the ASQ ("Have you ever tried to kill yourself?") and prefaced it with "In the past few weeks . . ." as they were not investigating lifetime suicidality. A response of 'no' was scored as 0 and a response of 'yes' was scored as 1; each item was summed, generating an overall score for suicidality on a scale ranging from 0 to 4, with higher scores indicating greater levels of suicidal ideation.
Beck Depression Inventory-II (BDI-II)	The BDI-II is a tool for assessing depressive symptoms. There are no specific scores to categorise depression severity, but it is suggested that 0 to 13 is minimal symptoms, 14 to 19 is mild depression, 20 to 28 is moderate depression, and severe depression is 29 to 63.
Body Image Scale (BIS)	The BIS is used to measure body satisfaction. The scale consists of 30 body features, which the person rates on a 5-point scale. Each of the 30 items falls into one of 3 basic groups based on its relative importance as a gender-defining body feature: primary sex characteristics, secondary sex characteristics, and neutral body characteristics. A higher score indicates more dissatisfaction.
Bone mineral apparent density (BMAD)	BMAD is a size adjusted value of bone mineral density (BMD) incorporating bone size measurements using UK norms in growing adolescents.
Center for Epidemiologic Studies Depression scale (CESD-R)	The CESD-R is a valid, widely used tool to access depressive symptoms. The CESD-R asks about how frequently a person has felt or behaved in a certain way; with 20 questions scored from 0 score is calculated as a sum of 20 questions, ranging from 0 ("not at all or less than one day") to 3 ("5–7 days" and/or "nearly every day for 2 weeks"). Total score ranges from 0 to 60, with higher scores indicating more depressive symptoms.
Cisgender	Cisgender is a term for someone whose gender identity matches their birth-registered sex.
Family APGAR (Adaptability, Partnership, Growth, Affection and Resolve) test	The Family APGAR test is a 5-item questionnaire, with higher scores indicating better family functioning. The authors reported the following interpretation of the score: functional, 17-20 points; mildly dysfunctional, 16-13 points; moderately dysfunctional, 12-10 point; severely dysfunctional, <9 points.
Gender	The roles, behaviours, activities, attributes and opportunities that any society considers appropriate for girls and boys, and women and men.
Gender dysphoria	Discomfort or distress that is caused by a discrepancy between a person's gender identity (how they see themselves regarding their gender) and that person's sex assigned at birth (and the associated gender role, and/or primary and secondary sex characteristics).

General Well-Being Scale (GWBS) of the Pediatric Quality of Life Inventory score	The GWBS of the Pediatric Quality of Life Inventory uses uses a 5-point response scale, contains seven items, and measures two dimensions: general wellbeing (6 items) and general health (1 item). Each item is scored from 0 to 4, and the total score is linearly transformed to a 0 to 100 scale. High scores reflect fewer perceived problems and greater well-being.
GnRH analogue	GnRH analogues competitively block GnRH receptors to prevent the spontaneous release of two gonadotropin hormones, Follicular Stimulating Hormone (FSH) and Luteinising Hormone (LH) from the pituitary gland. The reduction in LH and FSH secretion reduces oestradiol secretion from the ovaries in those whose sex assigned at birth was female and testosterone secretion from the testes in those whose sex assigned at birth was male.
Patient Health Questionnaire Modified for Teens score (PHQ 9_Modified for Teens)	The PHQ 9_Modified for Teens is a validated tool to assess depression, dysthymia and suicide risk. The tool consists of 9 questions scored from 0 to 3 (total score 0 to 27), plus an additional 4 questions that are not scored. A score of 0 to 4 suggests no or minimal depressive symptoms, 5 to 9 mild, 10-14 moderate, 15-19 moderate and 20-27 severe symptoms.
Quick Inventory of Depressive Symptoms (QIDS)	Both the clinician- and self-reported QIDS are validated tools to assess depressive symptoms. The tool consists of 16 items, with the highest score for 9 items (sleep, weight, psychomotor changes, depressed mood, decreased interest, fatigue, guilt, concentration, and suicidal ideation) are added to give a total score ranging from 0 to 27. A score of 0 to 5 is suggestive of no depressive symptoms, 6 to 10 mild symptoms, 11 to 15 moderate symptoms, 16-20 severe symptoms and 21 to 27 very severe symptoms.
Quality of Life Enjoyment and Satisfaction Questionnaire (QLES-Q-SF)	QLES-Q-SF is a validated questionnaire, consisting of 15 questions that rate quality of life on a scale of 1 (poor) to 5 (very good).
Screen for Child Anxiety Related Emotional Disorders (SCARED) questionnaire	SCARED is a validated, 41-point questionnaire, with each item scored 0 to 2. A total score of 25 or more is suggestive of anxiety disorder, with scores above 30 being more specific. Certain scores for specific questions may indicate the presence of other anxiety-related disorders: A score of 7 or more in questions related to panic disorder or significant somatic symptoms may indicate the presence of these. A score of 9 or more in questions related to generalised anxiety disorder may indicate the presence of this. A score of 5 or more in questions related to separation anxiety may indicate the presence of this. A score of 8 or more in questions related to social anxiety disorder may indicate the presence of this. A score of 3 or more in questions related to significant school avoidance may indicate the presence of this.
State-Trait Anxiety Inventory (STAI) score	STAI is a validated and commonly used measure of state anxiety (current state of anxiety) and trait anxiety (general state of calmness, confidence and security). It has 40 items, the first 20 covering state anxiety, the second 20 covering trait anxiety. STAI

	can be used in clinical settings to diagnose anxiety and to distinguish it from depressive illness. Each subtest (state and trait) is scored between 20 and 80, with higher scores indicating greater anxiety. There is no published minimal clinically meaningful difference (MCID) for STAI or thresholds for anxiety severity.
Strengths and Difficulties Questionnaire (SDQ, Spanish version	The SDQ, Spanish version includes 25-items covering emotional symptoms, conduct problems, hyperactivity/ inattention, peer relationship problems and prosocial behaviour. The authors state that a score of more than 20 is considered indicative of risk of having a disorder (normal: 0-15; borderline: 16-19, abnormal: 20-40).
Tanner stage	Tanner staging is a scale of physical development.
Transgender (including transmale and transfemale)	Transgender is a term for someone whose gender identity is not congruent with their birth-registered sex. A transfemale is a person who identifies as female and a transmale is a person who identifies as male.
Utrecht Gender Dysphoria Scale (UGDS)	The UGDS is a validated screening tool for both adolescents and adults to assess gender dysphoria. It consists of 12 items, to be answered on a 1- to 5-point scale, resulting in a sum score between 12 and 60. Higher scores indicate higher levels of gender dysphoria.

References

Included studies

- Achille, C., Taggart, T., Eaton, N.R. et al. (2020) Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: Preliminary results. International Journal of Pediatric Endocrinology 2020(1): 8
- Allen, LR, Watson, LB, Egan, AM et al. (2019) Well-being and suicidality among transgender youth after gender-affirming hormones. Clinical Practice in Pediatric Psychology 7(3): 302-311
- Kaltiala, R., Heino, E., Tyolajarvi, M. et al. (2020) Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria. Nordic Journal of Psychiatry 74(3): 213-219
- Khatchadourian K, Amed S, Metzger DL (2014) Clinical management of youth with gender dysphoria in Vancouver. The Journal of pediatrics 164(4): 906-11
- Klaver, Maartje, de Mutsert, Renee, van der Loos, Maria A T C et al. (2020) Hormonal Treatment and Cardiovascular Risk Profile in Transgender Adolescents. Pediatrics 145(3)
- Klink D, Caris M, Heijboer A et al. (2015) Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. The Journal of Clinical Endocrinology and Metabolism 100(2): e270-5

- Kuper, Laura E, Stewart, Sunita, Preston, Stephanie et al. (2020) Body Dissatisfaction and Mental Health Outcomes of Youth on Gender-Affirming Hormone Therapy. Pediatrics 145(4)
- Lopez de Lara, D., Perez Rodriguez, O., Cuellar Flores, I. et al. (2020) Psychosocial assessment in transgender adolescents. Anales de Pediatria
- Stoffers, Iris E; de Vries, Martine C; Hannema, Sabine E (2019) Physical changes, laboratory parameters, and bone mineral density during testosterone treatment in adolescents with gender dysphoria. The journal of sexual medicine 16(9): 1459-1468
- Vlot MC, Klink DT, den Heijer M et al. (2017) Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone 95: 11-19

Other references

- World Health Organisation (2018) International Classification of Diseases 11. Available from https://icd.who.int/ [accessed 27 August 2020]
- American Psychiatric Association. (2013). Diagnostic and statistical Manual of Mental Disorders (DSM-5) (5th ed). Washington, DC and London: American Psychiatric Publishing. pp.451-460. Available from: https://www.psychiatry.org/patients-families/gender-dysphoria/what-is-gender-dysphoria [accessed 27 August 2020]
- NHS England (2013). NHS Standard contract for gender identity development service for children and adolescents https://www.england.nhs.uk/wp-content/uploads/2017/04/gender-development-service-children-adolescents.pdf [accessed 27 August 2020]
- NHS England (2016). Clinical Commissioning Policy: Prescribing of Cross-Sex Hormones as part of the Gender Identity Development Service for Children and Adolescents https://www.england.nhs.uk/wp-content/uploads/2018/07/Prescribing-of-cross-sex-hormones-as-part-of-the-gender-identity-development-service-for-children-and-adolesce.pdf [accessed 27 August 2020]

Copyright

Evidence review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria

This document will help inform Dr Hilary Cass' independent review into gender identity services for children and young people. It was commissioned by NHS England and Improvement who commissioned the Cass review. It aims to assess the evidence for the clinical effectiveness, safety and cost-effectiveness of gonadotrophin releasing hormone (GnRH) analogues for children and adolescents aged 18 years or under with gender dysphoria.

The document was prepared by NICE in October 2020.

The content of this evidence review was up to date on 14 October 2020. See summaries of product characteristics (SPCs), British National Formulary (BNF) or the Medicines and Healthcare products Regulatory Agency (MHRA) or NICE websites for up-to-date information.

Contents

1. Introduction

This review aims to assess the evidence for the clinical effectiveness, safety and cost-effectiveness of gonadotrophin releasing hormone (GnRH) analogues for children and adolescents aged 18 years or under with gender dysphoria. The review follows the NHS England Specialised Commissioning process and template and is based on the criteria outlined in the PICO framework (see appendix A). This document will help inform Dr Hilary Cass' independent review into gender identity services for children and young people.

Gender dysphoria in children, also known as gender identity disorder or gender incongruence of childhood (World Health Organisation 2020), refers to discomfort or distress that is caused by a discrepancy between a person's gender identity (how they see themselves[1] regarding their gender) and that person's sex assigned at birth and the associated gender role, and/or primary and secondary sex characteristics (Diagnostic and Statistical Manual of Mental Disorders 2013).

GnRH analogues suppress puberty by delaying the development of secondary sexual characteristics. The intention is to alleviate the distress associated with the development of secondary sex characteristics, thereby providing a time for on-going discussion and exploration of gender identity before deciding whether to take less reversible steps. In England, the GnRH analogue triptorelin (a synthetic decapeptide analogue of natural GnRH, which has marketing authorisations for the treatment of prostate cancer, endometriosis and precocious puberty [onset before 8 years in girls and 10 years in boys]) is used for this purpose. The use of triptorelin for children and adolescents with gender dysphoria is off-label.

For children and adolescents with gender dysphoria it is recommended that management plans are tailored to the needs of the individual, and aim to ameliorate the potentially negative impact of gender dysphoria on general developmental processes, support young people and their families in managing the uncertainties inherent in gender identity development and provide on-going opportunities for exploration of gender identity. The plans may also include psychological support and exploration and, for some individuals, the use of GnRH analogues in adolescence to suppress puberty; this may be followed later with gender-affirming hormones of the desired sex (NHS England 2013).

2. Executive summary of the review

Nine observational studies were included in the evidence review. Five studies were retrospective observational studies (Brik et al. 2020, Joseph et al. 2019, Khatchadourian et al. 2014, Klink et al. 2015, Vlot et al. 2017), 3 studies were prospective longitudinal observational studies (Costa et al. 2015, de Vries et al. 2011, Schagen et al. 2016) and 1 study was a cross-sectional study (Staphorsius et al. 2015). Two studies (Costa et al. 2015

[1] Gender refers to the roles, behaviours, activities, attributes and opportunities that any society considers appropriate for girls and boys, and women and men (World Health Organisation, Health Topics: Gender).

and Staphorsius et al. 2015) provided comparative evidence and the remaining 7 studies used within-person, before and after comparisons.

The terminology used in this topic area is continually evolving and is different depending on stakeholder perspectives. In this evidence review we have used the phrase 'people's assigned sex at birth' rather than natal or biological sex, gonadotrophin releasing hormone (GnRH) analogues rather than 'puberty blockers' and gender-affirming hormones rather than 'cross sex hormones'. The research studies included in this evidence review may use historical terms which are no longer considered appropriate.

In children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Critical outcomes

The critical outcomes for decision making are the impact on gender dysphoria, mental health and quality of life. The quality of evidence for these outcomes was assessed as very low certainty using modified GRADE.

Impact on gender dysphoria
The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones does not affect gender dysphoria (measured using the Utrecht Gender Dysphoria Scale [UGDS]). The mean (±SD) gender dysphoria (UGDS) score was not statistically significantly different at baseline compared with follow-up (n=41, 53.20 [±7.91] versus 53.9 [±17.42], p=0.333).

Impact on mental health
The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones may reduce depression (measured using the Beck Depression Inventory-II [BDI-II]). The mean [±SD] BDI score was statistically significantly lower (improved) from baseline compared with follow-up (n=41, 8.31 [±7.12] versus 4.95 [±6.72], p=0.004).

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones does not affect anger (measured using the Trait Anger Scale [TPI]). The mean [±SD] anger (TPI) score was not statistically significantly different at baseline compared with follow-up (n=41, 18.29 [±5.54] versus 17.88 [±5.24], p=0.503).

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones does not affect anxiety (measured using the Trait Anxiety Scale [STAI]). The mean [±SD] anxiety (STAI) score was not statistically significantly different at baseline compared with follow-up (n=41, 39.43 [±10.07] versus 37.95 [±9.38], p=0.276).

Impact on quality of life
No evidence was identified.

Important outcomes

The important outcomes for decision making are impact on body image, psychosocial impact, engagement with health care services, impact on extent of and satisfaction with surgery and stopping treatment. The quality of evidence for all these outcomes was assessed as very low certainty using modified GRADE.

Impact on body image
The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones does not affect body image (measured using the Body Image Scale [BIS]). The mean [±SD] body image (BIS) scores were not statistically significantly different from baseline compared with follow-up for primary sexual characteristics (n=57, 4.10 [±0.56] versus 3.98 [±0.71], p=0.145), secondary sexual characteristics (n=57, 2.74 [±0.65] versus 2.82 [±0.68], p=0.569) or neutral body characteristics (n=57, 2.41 [±0.63] versus 2.47 [±0.56], p=0.620).

Psychosocial impact
The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that treatment with GnRH analogues before starting gender-affirming hormones may improve psychosocial impact over time (measured using the Children's Global Assessment Scale [CGAS]). The mean [±SD] CGAS score was statistically significantly higher (improved) from baseline compared with follow-up (n=41, 70.24 [±10.12] versus 73.90 [±9.63], p=0.005).

This study also found that psychosocial functioning may improve over time (measured using the Child Behaviour Checklist [CBCL] and the self-administered Youth Self-Report [YSR]). The mean [±SD] CBCL scores were statistically significantly lower (improved) from baseline compared with follow-up for Total T score (n=54, 60.70 [±12.76] versus 54.46 [±11.23], p<0.001), internalising T score (n=54, 61.00 [±12.21] versus 52.17 [±9.81], p<0.001) and externalising T score (n=54, 58.04 [±12.99] versus 53.81 [±11.86], p=0.001). The mean [±SD] YSR scores were statistically significantly lower (improved) from baseline compared with follow-up for Total T score (n=54, 55.46 [±11.56] versus 50.00 [±10.56], p<0.001), internalising T score (n=54, 56.04 [±12.49] versus 49.78 [±11.63], p<0.001) and externalising T score (n=54, 53.30 [±11.87] versus 49.98 [±9.35], p=0.009). The proportion of adolescents scoring in the clinical range decreased from baseline to follow up on the CBCL total problem scale (44.4% versus 22.2%, p=0.001) and the internalising scale of the YSR (29.6% versus 11.1%, p=0.017).

The study by Costa et al. 2015 in 201 adolescents with gender dysphoria who had 6 months of psychological support followed by either GnRH analogues and continued psychological support or continued psychological support only, found that during treatment with GnRH analogues psychosocial impact in terms of global functioning may improve over time (measured using the CGAS). In the group receiving GnRH analogues, the mean [±SD] CGAS score was statistically significantly higher (improved) after 6 months (n=60, 64.70 [±13.34]) and 12 months (n=35, 67.40 [±13.39]) compared with baseline (n=101, 58.72 [±11.38], p=0.003 and p<0.001, respectively). However, there was no statistically significant difference in global functioning (CGAS scores) between the group receiving GnRH analogues plus psychological support and the group receiving psychological support only at any time point.

The study by Staphorsius et al. 2015 in 40 adolescents with gender dysphoria (20 of whom were receiving GnRH analogues) gave mean [±SD] CBCL scores for each group, but statistical analysis is unclear (transfemales receiving GnRH analogues 57.4 [±9.8], transfemales not receiving GnRH analogues 58.2 [±9.3], transmales receiving GnRH analogues 57.5 [±9.4], transmales not receiving GnRH analogues 63.9 [±10.5]).

Engagement with health care services
The study by Brik et al. 2018 in 143 children and adolescents with gender dysphoria receiving GnRH analogues found that 9 adolescents in the original sampling frame (9/214, 4.2%) were excluded from the study because they stopped attending appointments.

The study by Costa et al. 2015 in 201 adolescents with gender dysphoria who had 6 months of psychological support followed by either GnRH analogues and continued psychological support or continued psychological support only had a large loss to follow-up over time. The sample size at baseline and 6 months was 201, which dropped by 39.8% to 121 after 12 months and by 64.7% to 71 at 18 months follow-up. No explanation of the reasons for loss to follow-up are reported.

Impact on extent of and satisfaction with surgery
No evidence was identified.

Stopping treatment
The study by Brik et al. 2018 in 143 children and adolescents with gender dysphoria receiving GnRH analogues reported the reasons for stopping GnRH analogues. During the follow-up period 6.2% (9/143) of adolescents had stopped GnRH analogues after a median duration of 0.8 years (range 0.1 to 3.0). Five adolescents stopped treatment because they no longer wished to receive gender-affirming treatment for various reasons. In 4 adolescents (all transmales), GnRH analogues were stopped mainly because of adverse effects (such as mood and emotional lability), although they wanted to continue treatments for gender dysphoria.

The study by Khatchadourian et al. 2014 in 27 adolescents with gender dysphoria who started GnRH analogues reported the reasons for stopping them. Eleven out of 26 where data was available (42%) stopped GnRH analogues during follow up.

In children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Evidence was available for bone density, cognitive development or functioning, and other safety outcomes. The quality of evidence for all these outcomes was assessed as very low certainty using modified GRADE.

Bone density
The study by Joseph et al. 2019 in 70 adolescents with gender dysphoria found that GnRH analogues may reduce the expected increase in lumbar or femoral bone density (measured with the z-score). However, the z-scores were largely within 1 standard deviation of normal,

and actual lumbar or femoral bone density values were not statistically significantly different between baseline and follow-up:

- The mean z-score [±SD] for lumbar bone mineral apparent density (BMAD) was statistically significantly lower at 1 year compared with baseline in transfemales (baseline 0.859 [±0.154], 1 year −0.228 [±1.027], p=0.000) and transmales (baseline −0.186 [±1.230], 1 year −0.541 [±1.396], p=0.006).
- The mean z-score [±SD] for lumbar BMAD was statistically significantly lower after receiving GnRH analogues for 2 years compared with baseline in transfemales (baseline 0.486 [±0.809], 2 years −0.279 [±0.930], p=0.000) and transmales (baseline −0.361 [±1.439], 2 years −0.913 [±1.318], p=0.001).
- The mean z-score [±SD] for femoral neck bone mineral density (BMD) was statistically significantly lower after receiving GnRH analogues for 2 years compared with baseline in transfemales (baseline 0.0450 [±0.781], 2 years −0.600 [±1.059], p=0.002) and transmales (baseline −1.075 [±1.145], 2 years −1.779 [±0.816], p=0.001).

The study by Klink et al. 2015 in 34 adolescents with gender dysphoria found that GnRH analogues may reduce the expected increase in lumbar (transmales only), but not femoral bone density. However, the z-scores are largely within 1 standard deviation of normal. Actual lumbar or femoral bone density values were not statistically significantly different between baseline and follow-up (apart from BMD measurements in transmales):

- The mean z-score [±SD] for lumbar BMAD was not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales, but was statistically significantly lower when starting gender-affirming hormones in transmales (GnRH analogues 0.28 [±0.90], gender-affirming hormones −0.50 [±0.81], p=0.004).

The study by Vlot et al. 2017 in 70 adolescents with gender dysphoria found that GnRH analogues may reduce the expected increase in lumbar or femoral bone density. However, the z-scores were largely within 1 standard deviation of normal. Actual lumbar or femoral bone density values were not statistically significantly different between baseline and follow-up (apart from in transmales with a bone age ≥14 years). This study reported change in bone density from starting GnRH analogues to starting gender-affirming hormones by bone age:

- The median z-score [range] for lumbar BMAD in transfemales with a bone age of <15 years was statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (GnRH analogues −0.20 [−1.82 to 1.18], gender-affirming hormones −1.52 [−2.36 to 0.42], p=0.001) but was not statistically significantly different in transfemales with a bone age ≥15 years.
- The median z-score [range] for lumbar BMAD in transmales with a bone age of <14 years was statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (GnRH analogues −0.05 [−0.78 to 2.94], gender-affirming hormones −0.84 [−2.20 to 0.87], p=0.003) and in transmales with a bone age ≥14 years (GnRH analogues 0.27 [−1.60 to 1.80], gender-affirming hormones −0.29 [−2.28 to 0.90], p≤0.0001).

- The median z-score [range] for femoral neck BMAD in transfemales with a bone age of <15 years was not statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (GnRH analogues −0.71 [−3.35 to 0.37], gender-affirming hormones −1.32 [−3.39 to 0.21], p≤0.1) or in transfemales with a bone age ≥15 years (GnRH analogues −0.44 [−1.37 to 0.93], gender-affirming hormones −0.36 [−1.50 to 0.46]).
- The z-score for femoral neck BMAD in transmales with a bone age of <14 years was not statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (GnRH analogues −0.01 [−1.30 to 0.91], gender-affirming hormone −0.37 [−2.28 to 0.47]) but was statistically significantly lower in transmales with a bone age ≥14 years (GnRH analogues 0.27 [−1.39 to 1.32], gender-affirming hormones −0.27 [−1.91 to 1.29], p=0.002).

Cognitive development or functioning

The study by Staphorsius et al. 2015 in 40 adolescents with gender dysphoria (20 of whom were receiving GnRH analogues) measured cognitive development or functioning (using an IQ test, and reaction time and accuracy measured using the Tower of London task):

- The mean (±SD) IQ in transfemales receiving GnRH analogues was 94.0 (±10.3) and 109.4 (±21.2) in the control group. In transmales receiving GnRH analogues the mean (±SD) IQ was 95.8 (±15.6) and 98.5 (±15.9) in the control group.
- The mean (±SD) reaction time in transfemales receiving GnRH analogues was 10.9 (±4.1) and 9.9 (±3.1) in the control group. In transmales receiving GnRH analogue it was 9.9 (±3.1) and 10.0 (±2.0) in the control group.
- The mean (±SD) accuracy score in transfemales receiving GnRH analogues was 73.9 (±9.1) and 83.4 (±9.5) in the control group. In transmales receiving GnRH analogues it was 85.7 (±10.5) and 88.8 (±9.7) in the control group.

No statistical analyses or interpretation of the results was reported.

Other safety outcomes

The study by Schagen et al. 2016 in 116 adolescents with gender dysphoria found that GnRH analogues do not affect renal or liver function:

- There was no statistically significant difference between baseline and 1 year results for serum creatinine in transfemales, but there was a statistically significant decrease between baseline and 1 year in transmales (p=0.01).
- Glutamyl transferase, alanine aminotransferase (ALT), and aspartate aminotransferase (AST) levels did not significantly change from baseline to 12 months of treatment.

The study by Khatchadourian et al. 2014 in 27 adolescents with gender dysphoria who started GnRH analogues narratively reported adverse effects from GnRH analogues in 26 adolescents:

- 1 transmale developed sterile abscesses; they were switched from leuprolide acetate to triptorelin, and this was well tolerated
- 1 transmale developed leg pains and headaches, which eventually resolved
- 1 participant gained 19 kg within 9 months of starting GnRH analogues.

In children and adolescents with gender dysphoria, what is the cost-effectiveness of GnRH analogues compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?

No cost-effectiveness evidence was found for GnRH analogues in children and adolescents with gender dysphoria.

From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may benefit from GnRH analogues more than the wider population of interest?

Some studies reported data separately for the following subgroups of children and adolescents with gender dysphoria: sex assigned at birth males (transfemales) and sex assigned at birth females (transmales). This included some direct comparisons of these subgroups, and differences were largely seen at baseline as well as follow up. No evidence was found for other specified subgroups.

Sex assigned at birth males (transfemales)
Impact on gender dysphoria
The study by Costa et al. 2015 in 201 adolescents with gender dysphoria who had 6 months of psychological support followed by either GnRH analogues and continued psychological support or continued psychological support only, found that gender dysphoria (measured using the UGDS) in sex assigned at birth males is lower than in sex assigned at birth females. Sex assigned at birth males had a statistically significantly lower (improved) mean [±SD] UGDS score of 51.6 [±9.7] compared with sex assigned at birth females (56.1 [±4.3], p<0.001), but it was not reported if this was at baseline or follow-up.

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that gender dysphoria (measured using the UGDS) in sex assigned at birth males is lower than in sex assigned at birth females at baseline and follow up. The mean [±SD] UGDS score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean UGDS score: 47.95 [±9.70] versus 56.57 [±3.89]) and follow up (n=not reported, 49.67 [±9.47] versus 56.62 [±4.00]); between sex difference p<0.001).

Impact on mental health
The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that the impact on mental health (depression, anger and anxiety) may be different in sex assigned at birth males compared with sex assigned at birth females. Over time there was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for depression, but sex assigned at birth males had statistically significantly lower levels of anger and anxiety than sex assigned at birth females at baseline and follow up.

- The mean [±SD] depression (BDI-II) score was not statistically significantly different in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean BDI score [±SD]: 5.71 [±4.31] versus 10.34 [±8.24]) and follow-up (n=not reported, 3.50 [±4.58] versus 6.09 [±7.93]), between sex difference p=0.057

- The mean [±SD] anger (TPI) score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean TPI score [±SD]: 5.22 [±2.76] versus 6.43 [±2.78]) and follow-up (n=not reported, 5.00 [±3.07] versus 6.39 [±2.59]), between sex difference p=0.022
- The mean [±SD] anxiety (STAI) score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean STAI score [±SD]: 4.33 [±2.68] versus 7.00 [±2.36]) and follow-up (n=not reported, 4.39 [±2.64] versus 6.17 [±2.69]), between sex difference p<0.001.

Impact on body image

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that the impact on body image may be different in sex assigned at birth males compared with sex assigned at birth females. Sex assigned at birth males are less dissatisfied with their primary and secondary sex characteristics than sex assigned at birth females at both baseline and follow up, but the satisfaction with neutral body characteristics is not different.

- The mean [±SD] BIS score for primary sex characteristics was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean BIS score [±SD]: 4.02 [±0.61] versus 4.16 [±0.52]) and follow up (n=not reported, 3.74 [±0.78] versus 4.17 [±0.58]) between sex difference p=0.047.
- The mean [±SD] BIS score for secondary sex was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, mean BIS score [±SD]: 2.66 [±0.50] versus 2.81 [±0.76]) and follow up (n=not reported, 2.39 [±0.69] versus 3.18 [±0.42]), between sex difference p=0.001.
- The mean [±SD] BIS score for neutral body characteristics was not statistically significantly different in sex assigned at birth males compared with sex assigned at birth females at baseline (n=not reported, 2.60 [±0.58] versus 2.24 [±0.62], between sex difference p=0.777).

Psychosocial impact

The study by Costa et al. 2015 in 201 adolescents with gender dysphoria who had 6 months of psychological support followed by either GnRH analogues and continued psychological support or continued psychological support only, found that sex assigned at birth males had statistically significant lower mean [±SD] CGAS scores at baseline compared with sex assigned at birth females (n=201, 55.4 [±12.7] versus 59.2 [±11.8], p=0.03), but no conclusions could be drawn.

The study by de Vries et al. 2011 in 70 adolescents with gender dysphoria found that psychosocial impact in terms of global functioning (CGAS) and psychosocial functioning (CBCL and YSR) may be different in sex assigned at birth males compared with sex assigned at birth females, but no conclusions could be drawn.

- There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females (at baseline or follow up) for the CBCL Total T

score, the CBCL internalising T score, the YSR Total T score or the YSR internalising T score.

- Sex assigned at birth males had statistically higher mean [±SD] CGAS scores compared with sex assigned at birth females at baseline (n=54, 73.10 [±8.44] versus 67.25 [±11.06]) and follow up (n=54, 77.33 [±8.69] versus 70.30 [±9.44]), between sex difference p=0.021.
- Sex assigned at birth males had statistically lower mean [±SD] CBCL externalising T scores compared with sex assigned at birth females at baseline (n=54, 54.71 [±12.91] versus 60.70 [±12.64]) and follow up (n=54, 48.75 [±10.22] versus 57.87 [±11.66]), between sex difference p=0.015.
- Sex assigned at birth males had statistically lower mean [±SD] YSR externalising T scores compared with sex assigned at birth females at both baseline (n=54, 48.72 [±11.38] versus 57.24 [±10.59]) and follow up (n=54, 46.52 [±9.23] versus 52.97 [±8.51]), between sex difference p=0.004.

Bone density

The studies by Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017 provided evidence on bone density in sex assigned at birth males (see above for details).

Cognitive development or functioning

The study by Staphorsius et al. 2015 provided evidence on cognitive development or functioning in sex assigned at birth males (see above for details).

Other safety outcomes

The study by Schagen et al. 2016 provided evidence on renal function in sex assigned at birth males (see above).

Sex assigned at birth females (transmales)

Impact on gender dysphoria

The studies by de Vries et al. 2011 and Costa et al. 2015 found that gender dysphoria (measured using the UGDS) in sex assigned at birth females is higher than in sex assigned at birth males at baseline and follow up (see above for details).

Impact on mental health

The study by de Vries et al. 2011 found that the impact on mental health (depression, anger and anxiety) may be different in sex assigned at birth females compared with sex assigned at birth males. Over time there was no statistically significant difference between sex assigned at birth females and sex assigned at birth males for depression, but sex assigned at birth females had statistically significantly greater levels of anger and anxiety than sex assigned at birth males at both baseline and follow up (see above for details).

Impact on body image

The study by de Vries et al. 2011 found that the impact on body image may be different in sex assigned at birth females compared with sex assigned at birth males. Sex assigned at birth females are more dissatisfied with their primary and secondary sex characteristics than sex assigned at birth males at both baseline and follow up, but the satisfaction with neutral body characteristics is not different (see above for details).

Psychosocial impact
The studies by de Vries et al. 2011 and Costa et al. 2015 found that psychosocial impact in terms of global functioning (CGAS) and psychosocial functioning (CBCL and YSR) may be different in sex assigned at birth females compared with sex assigned at birth males, but no conclusions could be drawn (see above for details).

Bone density
The studies by Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017 provided evidence on bone density in sex assigned at birth females (see above for details).

Cognitive development or functioning
The study by Staphorsius et al. 2015 provided evidence on cognitive development or functioning in sex assigned at birth females (see above for details).

Other safety outcomes
The study by Schagen et al. 2016 provided evidence on renal function in sex assigned at birth females (see above for details).

From the evidence selected:

(a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?

(b) what were the ages at which participants commenced treatment with GnRH analogues?

(c) what was the duration of treatment with GnRH analogues?

All studies that reported diagnostic criteria for gender dysphoria (6/9 studies) used the version of the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria that was in use at the time. In 5 studies (Costa et al. 2015, Klink et al. 2015, Schagen et al. 2016, Staphorsius et al. 2015 and Vlot et al. 2017) the DSM-fourth edition, text revision (IV-TR) criteria were used. The study by Brik et al. 2020 used DSM-V criteria. It was not reported how gender dysphoria was defined in the remaining 3 studies.

The studies show variation in the age (11 to 18 years old) at which children and adolescents with gender dysphoria started GnRH analogues.

Most studies did not report the duration of treatment with GnRH analogues (Joseph et al. 2019, Khatchadourian et al. 2014, Vlot et al. 2017, Costa et al. 2015, de Vries et al. 2011, Schagen et al. 2016), but where this was reported (Brik et al. 2020, Klink et al. 2015, Staphorsius et al. 2015) there was a wide variation ranging from a few months to about 5 years.

Discussion

A key limitation to identifying the effectiveness and safety of GnRH analogues for children and adolescents with gender dysphoria is the lack of reliable comparative studies. The lack of clear, expected outcomes from treatment with a GnRH analogue (the purpose of which is to suppress secondary sexual characteristics which may cause distress from unwanted pubertal changes) also makes interpreting the evidence difficult.

The studies included in this evidence review are all small, uncontrolled observational studies, which are subject to bias and confounding, and all the results are of very low certainty using modified GRADE. They all reported physical and mental health comorbidities and concomitant treatments very poorly. All the studies are from a limited number of, mainly European, care facilities. They are described as either tertiary referral or expert services but the low number of services providing such care and publishing evidence may bias the results towards the outcomes in these services only and limit extrapolation.

Many of the studies did not report statistical significance or confidence intervals. Changes in outcome scores for clinical effectiveness and bone density were assessed with regards to statistical significance. However, there is relatively little interpretation of whether the changes in outcomes are clinically meaningful.

In the observational, retrospective studies providing evidence on bone density, participants acted as their own controls and change in bone density was determined between starting GnRH analogues and follow up. Observational studies such as these can only show an association with GnRH analogues and bone density; they cannot show that GnRH analogues caused any differences in bone density seen. Because there was no comparator group and participants acted as their own controls, it is not known whether the findings are associated with GnRH analogues or due to changes over time.

Conclusion

The results of the studies that reported impact on the critical outcomes of gender dysphoria and mental health (depression, anger and anxiety), and the important outcomes of body image and psychosocial impact (global and psychosocial functioning), in children and adolescents with gender dysphoria are of very low certainty using modified GRADE. They suggest little change with GnRH analogues from baseline to follow-up.

Studies that found differences in outcomes could represent changes that are either of questionable clinical value, or the studies themselves are not reliable and changes could be due to confounding, bias or chance. It is plausible, however, that a lack of difference in scores from baseline to follow-up is the effect of GnRH analogues in children and adolescents with gender dysphoria, in whom the development of secondary sexual characteristics might be expected to be associated with an increased impact on gender dysphoria, depression, anxiety, anger and distress over time without treatment. The study by de Vries et al. 2011 reported statistically significant reductions in the Child Behaviour Checklist (CBCL) and Youth Self-Report (YSR) scores from baseline to follow up, which include measures of distress. As the aim of GnRH analogues is to reduce distress caused by the development of secondary sexual characteristics, this may be an important finding. However, as the studies all lack appropriate controls who were not receiving GnRH analogues, any positive changes could be a regression to mean.

The results of the studies that reported bone density outcomes suggest that GnRH analogues may reduce the expected increase in bone density (which is expected during puberty). However, as the studies themselves are not reliable, the results could be due to confounding, bias or chance. While controlled trials may not be possible, comparative studies are needed to understand this association and whether the effects of GnRH analogues on bone density are seen after they are stopped. All the studies that reported safety outcomes provided very low certainty evidence.

No cost-effectiveness evidence was found to determine whether or not GnRH analogues are cost-effective for children and adolescents with gender dysphoria.

The results of the studies that reported outcomes for subgroups of children and adolescents with gender dysphoria, suggest there may be differences between sex assigned at birth males (transfemales) and sex assigned at birth females (transmales).

3. Methodology

Review questions

The review question(s) for this evidence review are:

1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
2. For children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
3. For children and adolescents with gender dysphoria, what is the cost-effectiveness of GnRH analogues compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?
4. From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may derive more (or less) advantage from treatment with GnRH analogues than the wider population of children and adolescents with gender dysphoria?
5. From the evidence selected,
   a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?
   b) what were the ages at which participants commenced treatment with GnRH analogues?
   c) what was the duration of treatment with GnRH analogues?

See appendix A for the full review protocol.

Review process

The methodology to undertake this review is specified by NHS England in their 'Guidance on conducting evidence reviews for Specialised Services Commissioning Products' (2020).

The searches for evidence were informed by the PICO document and were conducted on 23 July 2020.

See appendix B for details of the search strategy.

Results from the literature searches were screened using their titles and abstracts for relevance against the criteria in the PICO framework. Full text references of potentially

relevant evidence were obtained and reviewed to determine whether they met the inclusion criteria for this evidence review.

See appendix C for evidence selection details and appendix D for the list of studies excluded from the review and the reasons for their exclusion.

Relevant details and outcomes were extracted from the included studies and were critically appraised using a checklist appropriate to the study design. See appendices E and F for individual study and checklist details.

The available evidence was assessed by outcome for certainty using modified GRADE. See appendix G for GRADE Profiles.

4. Summary of included studies

Nine observational studies were identified for inclusion. Five studies were retrospective observational studies (Brik et al. 2020, Joseph et al. 2019, Khatchadourian et al. 2014, Klink et al. 2015, Vlot et al. 2017), 3 studies were prospective longitudinal observational studies (Costa et al. 2015, de Vries et al. 2011, Schagen et al. 2016) and 1 study was a cross-sectional study (Staphorsius et al. 2015).

The terminology used in this topic area is continually evolving and is different depending on stakeholder perspectives. In this evidence review we have used the phrase 'people's assigned sex at birth' rather than natal or biological sex, gonadotrophin releasing hormone (GnRH) analogues rather than 'puberty blockers' and gender-affirming hormones rather than 'cross sex hormones'. The research studies included in this evidence review may use historical terms which are no longer considered appropriate.

Table 1 provides a summary of these included studies and full details are given in appendix E.

Table 1 Summary of included studies

Study	Population	Intervention and comparison	Outcomes reported
Brik et al. 2020 Retrospective observational single-centre study Netherlands	The study was conducted at the Curium-Leiden University Medical Centre gender clinic in Leiden, the Netherlands and involved adolescents with gender dysphoria. The sample size was 143 adolescents (median age at start of treatment was 15.0 years, range 11.1 to 18.6 years in transfemales; 16.1 years, range 10.1 to 17.9 years in transmales) from a sampling frame of 269 children and adolescents registered at the clinic between November 2010 and January 2018.	**Intervention** 143 children and adolescents receiving GnRH analogues (no specific treatment, dose, route or frequency of administration reported). The median duration was 2.1 years (range 1.6–2.8 years). **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Stopping treatment

Study	Population	Intervention and comparison	Outcomes reported
	Participants were included in the study if they were diagnosed with gender dysphoria according to the DSM-5 criteria, registered at the clinic, were prepubertal and within the appropriate age range, and had started GnRH analogues. No concomitant treatments were reported.		
Costa et al. 2015 Prospective longitudinal observational single centre cohort study United Kingdom	The study was conducted at the Gender Identity Development Service in London and involved adolescents with gender dysphoria. The sample size was 201 adolescents (mean [±SD] age 15.52±1.41 years, range 12 to 17 years) from a sampling frame of 436 consecutive adolescents referred to the service between 2010 and 2014. The mean [±SD] age at the start of GnRH analogues was 16.48 [±1.26] years, range 13 to 17 years. Participants were invited to participate following a 6-month diagnostic process using DSM-IV-TR criteria. No concomitant treatments were reported.	**Intervention** 101 adolescents assessed as being immediately eligible for GnRH analogues (no specific treatment, dose or route of administration reported) plus psychological support. The average duration of treatment was approximately 12 months (no exact figure given). **Comparison** 100 adolescents assessed as not immediately eligible for GnRH analogues (more time needed to make the decision to start GnRH analogues) who had psychological support only. None received GnRH analogues throughout the study.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Psychosocial impact
de Vries et al. 2011 Prospective longitudinal observational single centre before and after study Netherlands	The study was conducted at the Amsterdam gender identity clinic of the VU University Medical Centre and involved adolescents who were defined as “transsexual”. The sample size was 70 adolescents receiving GnRH analogues (mean age [±SD] at assessment 13.6±1.8 years) from a sampling frame of 196 consecutive adolescents referred to the service between 2000 and 2008. Participants were invited to participate if they subsequently started gender-affirming hormones between 2003 and 2009. No diagnostic criteria or concomitant treatments were reported.	**Intervention** 70 individuals assessed at baseline (T0) before the start of GnRH analogues (no specific treatment, dose or route of administration reported). **Comparison** No comparator.	**Critical Outcomes** • Gender dysphoria • Mental health (depression, anger and anxiety) **Important outcomes** • Body image • Psychosocial impact

Study	Population	Intervention and comparison	Outcomes reported
Joseph et al. 2019 Retrospective longitudinal observational single centre study United Kingdom	This study was conducted at the Early intervention clinic at University College London Hospital (all participants had been seen at the Gender Identity Development Service in London) and involved adolescents with gender dysphoria. The sample size was 70 adolescents with gender dysphoria (no diagnostic criteria described) all offered GnRH analogues. The mean age at the start of treatment was 13.2 years (SD ±1.4) for transfemales and 12.6 years (SD ±1.0) for transmales. Details of the sampling frame were not reported. Further details of how the sample was drawn are not reported. No concomitant treatments were reported.	**Intervention** GnRH analogues. No specific treatment, duration, dose or route of administration reported. **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Safety: bone density
Khatchadourian et al. 2014 Retrospective observational chart review single centre study Canada	This study was conducted at the Endocrinology and Diabetes Unit at British Columbia Children's Hospital, Canada and involved youths with gender dysphoria. The sample size was 27 young people with gender dysphoria who started GnRH analogues (at mean age 14.7 [SD ±1.9] years) out of 84 young people seen at the unit between 1998 and 2011. Diagnostic criteria and concomitant treatments were not reported.	**Intervention** 84 young people with gender dysphoria. For GnRH analogues no specific treatment, duration, dose or route of administration reported. **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Stopping treatment • Safety: adverse effects
Klink et al. 2015 Retrospective longitudinal observational single centre study Netherlands	This study was conducted in the Netherlands at a tertiary referral centre. It is unclear which centre this was. The sample size was 34 adolescents (mean age 14.9 [SD ±1.9] years for transfemales and 15.0 [SD ±2.0] years for transmales at start of GnRH analogues). Details of the sampling frame are not reported. Participants were included if they met DSM-IV-TR criteria for gender identity disorder of adolescence and had been treated with GnRH analogues and gender-affirming hormones during their pubertal years. No concomitant treatments were reported.	**Intervention** The intervention was GnRH analogue monotherapy (triptorelin 3.75 mg subcutaneously every 4 weeks) followed by gender-affirming hormones with discontinuation of GnRH analogues after gonadectomy. Duration of GnRH analogues was 1.3 years (range 0.5 to 3.8 years) in transfemales and 1.5 years (0.25 to 5.2 years in transmales. **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Safety: bone density

Study	Population	Intervention and comparison	Outcomes reported
Schagen et al. 2016 Prospective longitudinal study Netherlands	This study was conducted at the Centre of Expertise on Gender Dysphoria at the VU University Medical Centre (Amsterdam, Netherlands) and involved adolescents with gender dysphoria. The sample size was 116 adolescents (median age [range] 13.6 years [11.6 to 17.9] in transfemales and 14.2 years [11.1 to 18.6] in transmales during first year of GnRH analogues) out of 128 adolescents who started GnRH analogues. Participants were included if they met DSM-IV-TR criteria for gender dysphoria, had lifelong extreme gender dysphoria, were psychologically stable and were living in a supportive environment. No concomitant treatments were reported.	**Intervention** The intervention was GnRH analogue monotherapy (triptorelin 3.75 mg at 0, 2 and 4 weeks followed by intramuscular injections every 4 weeks, for at least 3 months). **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Safety: liver and renal function.
Staphorsius et al. 2015 Cross-sectional (single time point) assessment single centre study Netherlands	This study was conducted at the VU University Medical Centre (Amsterdam, Netherlands) and involved adolescents with gender dysphoria. The sample size was 85, of whom 40 were adolescents with gender dysphoria (20 of whom were being treated with GnRH analogues) and 45 were controls without gender dysphoria (not further reported here). Mean (±SD) age 15.1 (±2.4) years in transfemales and 15.8 (±1.9) years in transmales. Details of the sampling frame are not reported. Participants were included if they were diagnosed with Gender Identity Disorder according to the DSM-IV-TR and at least 12 years old and Tanner stage of at least B2 or G2 to G3 with measurable oestradiol and testosterone levels in girls and boys, respectively. No concomitant treatments were reported.	**Intervention** The intervention was a GnRH analogue (triptorelin 3.75 mg every 4 weeks subcutaneously or intramuscularly). The mean duration of treatment was 1.6 years (SD ±1.0). **Comparison** Adolescents with gender dysphoria not treated with GnRH analogues.	**Critical Outcomes** • No critical outcomes reported **Important outcomes** • Psychosocial impact • Safety: cognitive functioning
Vlot et al. 2017 Retrospective observational data analysis study	This study was conducted at the VU University Medical Centre (Amsterdam, Netherlands) and involved adolescents with gender dysphoria. The sample size was 70 adolescents (median age [range] 15.1 years [11.7 to 18.6] for	**Intervention** The intervention was a GnRH analogue (triptorelin 3.75 mg every 4 weeks subcutaneously). **Comparison** No comparator.	**Critical Outcomes** • No critical outcomes reported **Important outcomes**

Study	Population	Intervention and comparison	Outcomes reported
Netherlands	transmales and 13.5 years [11.5 to 18.3] for transfemales at start of GnRH analogues). Details of the sampling frame are not reported. Participants were included if they had a diagnosis of gender dysphoria according to DSM-IV-TR criteria who were receiving GnRH analogues and then gender-affirming hormones. No concomitant treatments were reported.		• Safety: bone density
Abbreviations: DSM-IV-TR, Diagnostic and Statistical Manual of Mental Disorders, fourth edition, text revision; GnRH, Gonadotrophin releasing hormone; SD, Standard deviation.			

5. Results

In children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement
Clinical Effectiveness	
Critical outcomes	
Impact on gender dysphoria **Certainty of evidence: very low**	This is a critical outcome because gender dysphoria in children and adolescents is associated with significant distress and problems with functioning. One uncontrolled, prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on gender dysphoria in adolescents, measured using the Utrecht Gender Dysphoria Scale (UGDS). The UGDS is a validated screening tool for both adolescents and adults to assess gender dysphoria. It consists of 12 items, to be answered on a 1- to 5-point scale, resulting in a sum score between 12 and 60. The higher the UGDS score the greater the gender dysphoria. The study measured the impact on gender dysphoria at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) UGDS score was not statistically significantly different at baseline compared with follow-up (n=41, 53.20 [±7.91] versus 53.9 [±17.42], p=0.333) **(VERY LOW)**.

	This study provides very low certainty evidence that treatment with GnRH analogues, before starting gender-affirming hormones, does not affect gender dysphoria.
Impact on mental health: depression **Certainty of evidence: very low**	This is a critical outcome because self-harm and thoughts of suicide have the potential to result in significant physical harm and, for completed suicides, the death of the young person. One uncontrolled, prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on depression in children and adolescents with gender dysphoria. Depression was measured using the Beck Depression Inventory-II (BDI-II). The BDI-II is a valid, reliable, and widely used tool for assessing depressive symptoms. There are no specific scores to categorise depression severity, but it is suggested that 0 to 13 is minimal symptoms, 14 to 19 is mild depression, 20 to 28 is moderate depression, and severe depression is 29 to 63. The study provided evidence for depression measured at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) depression (BDI) score was statistically significantly lower (improved) from baseline compared with follow-up (n=41, 8.31 [±7.12] versus 4.95 [±6.72], p=0.004) **(VERY LOW)**. **This study provides very low certainty evidence that treatment with GnRH analogues, before starting gender-affirming hormones, may reduce depression.**
Impact on mental health: anger **Certainty of evidence: very low**	This is a critical outcome because self-harm and thoughts of suicide have the potential to result in significant physical harm and, for completed suicides, the death of the young person. One uncontrolled, prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on anger in children and adolescents with gender dysphoria. Anger was measured using the Trait Anger Scale of the State-Trait Personality Inventory (TPI). This is a validated 20-item inventory tool which measures the intensity of anger as the disposition to experience angry feelings as a personality trait. Higher scores indicate greater anger. The study provided evidence for anger measured at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) anger (TPI) score was not statistically significantly different at baseline compared with follow-up (n=41, 18.29 [±5.54] versus 17.88 [±5.24], p=0.503) **(VERY LOW)**. **This study provides very low certainty evidence that treatment with GnRH analogues, before starting gender-affirming hormones, does not affect anger.**

Impact on mental health: anxiety **Certainty of evidence: very low**	This is a critical outcome because self-harm and thoughts of suicide have the potential to result in significant physical harm and, for completed suicides, the death of the young person. One uncontrolled, prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on anxiety in children and adolescents with gender dysphoria. Anxiety was measured using the Trait Anxiety Scale of the State-Trait Personality Inventory (STAI). This is a validated and commonly used measure of trait and state anxiety. It has 20 items and can be used in clinical settings to diagnose anxiety and to distinguish it from depressive illness. Higher scores indicate greater anxiety. The study provided evidence for anxiety at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) anxiety (STAI) score was not statistically significantly different at baseline compared with follow-up (n=41, 39.43 [±10.07] versus 37.95 [±9.38], p=0.276) (**VERY LOW)**. **This study provides very low certainty evidence that treatment with GnRH analogues, before starting gender-affirming hormones, does not affect levels of anxiety.**
Quality of life	This is a critical outcome because gender dysphoria in children and adolescents may be associated with a significant reduction in health-related quality of life. No evidence was identified.
Important outcomes	
Impact on body image **Certainty of evidence: very low**	This is an important outcome because some children and adolescents with gender dysphoria may want to take steps to suppress features of their physical appearance associated with their sex assigned at birth or accentuate physical features of their desired gender. One uncontrolled, prospective observational longitudinal study provided evidence relating to the impact on body image (de Vries et al. 2011). Body image was measured using the Body Image Scale (BIS) which is a validated 30-item scale covering 3 aspects: primary, secondary and neutral body characteristics. Higher scores represent a higher degree of body dissatisfaction. The study (de Vries et al. 2011) provided evidence for body image measured at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) body image (BIS) scores for were not statistically significantly different from baseline compared with follow-up for:

	• primary sexual characteristics (n=57, 4.10 [±0.56] versus 3.98 [±0.71], p=0.145) • secondary sexual characteristics (n=57, 2.74 [±0.65] versus 2.82 [±0.68], p=0.569) • neutral body characteristics (n=57, 2.41 [±0.63] versus 2.47 [±0.56], p=0.620) **(VERY LOW)**. **This study provides very low certainty evidence that treatment with GnRH analogues, before starting gender affirming hormones, does not affect body image.**
Psychosocial impact: global functioning **Certainty of evidence: very low**	This is an important outcome because gender dysphoria in children and adolescents is associated with internalising and externalising behaviours, and emotional and behavioural problems which may impact on social and occupational functioning. One uncontrolled, observational, prospective cohort study (de Vries et al 2011) and one prospective cross-sectional cohort study (Costa et al. 2015) provided evidence relating to psychosocial impact in terms of global functioning. Global functioning was measured using the Children's Global Assessment Scale (CGAS). The CGAS tool is a validated measure of global functioning on a single rating scale from 1 to 100. Lower scores indicate poorer functioning. One study (de Vries et al. 2011) provided evidence for global functioning (CGAS) at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and • shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). The mean (±SD) CGAS score was statistically significantly higher (improved) from baseline compared with follow-up (n=41, 70.24 [±10.12] versus 73.90 [±9.63], p=0.005) **(VERY LOW)**. One study (Costa et al. 2015) in adolescents with gender dysphoria who had 6 months of psychological support followed by either GnRH analogues and continued psychological support (the immediately eligible group) or continued psychological support only (the delayed eligible group who did not receive GnRH analogues) provided evidence for global functioning (CGAS) measured at 4 time points: • at baseline (T0) in both groups, • after 6 months of psychological support in both groups (T1), • after 6 months of GnRH analogues and 12 months of psychological support in the immediately eligible group and 12 months of psychological support only in the delayed eligible group (T2), and • after 18 months of psychological support and 12 months of GnRH analogues in the immediately eligible group and after 18 months of psychological support only in the delayed eligible group (T3). The mean [±SD] CGAS score was statistically significantly higher (improved) for all adolescents (including those not receiving GnRH analogues) at T1, T2 or T3 compared with baseline (T0).

	For the immediately eligible group (who received GnRH analogues) versus the delayed eligible group (who did not receive GnRH analogues) there were no statistically significant differences in CGAS scores between the 2 groups at baseline T0 (n=201, p=0.23), T1 (n=201, p=0.73), T2 (n=121, p=0.49) or T3 (n=71, p=0.14) time points. For the immediately eligible group (who received GnRH analogues), the mean (±SD) CGAS score was not statistically significantly different at: • T1 compared with T0 • T2 compared with T1 • T3 compared with T2. The mean (±SD) CGAS score was statistically significantly higher (improved) at: • T2 compared with T0 (n=60, 64.70 [±13.34] versus n=101, 58.72 [±11.38], p=0.003) • T3 compared with T0 (n=35, 67.40 [±13.39] versus n=101, 58.72 [±11.38], p<0.001) • T3 compared with T1 (n=35, 67.40 [±13.93] versus n=101, 60.89 [±12.17], p<0.001) **(VERY LOW)**. **These studies provide very low certainty evidence that during treatment with GnRH analogues, global functioning may improve over time. However, there was no statistically significant difference in global functioning between GnRH analogues plus psychological support compared with psychological support only at any time point.**
Psychosocial impact: psychosocial functioning **Certainty of evidence: very low**	This is an important outcome because gender dysphoria in children and adolescents is associated with internalising and externalising behaviours, and emotional and behavioural problems which may impact on social and occupational functioning. Two studies provided evidence for this outcome. One uncontrolled, observational, prospective cohort study (de Vries et al, 2011) and 1 cross-sectional observational study (Staphorsius et al. 2015) assessed psychosocial functioning using the Child Behaviour Checklist (CBCL) and the self-administered Youth Self-Report (YSR). The CBCL is a checklist parents complete to detect emotional and behavioural problems in children and adolescents. YSR is similar but is self-completed by the child or adolescent. The scales consist of a Total problems score, which is the sum of the scores of all the problem items. An internalising problem scale sums the anxious/depressed, withdrawn-depressed, and somatic complaints scores while the externalising problem scale combines rule-breaking and aggressive behaviour. The standard scores are scaled so that 50 is average for the child or adolescent's age and gender, with a SD of 10 points. Higher scores indicate greater problems, with a T-score above 63 considered to be in the clinical range. One study (de Vries et al. 2011) provided evidence for psychosocial functioning (CBCL and YSR scores) at 2 time points: • before starting a GnRH analogue (mean [±SD] age: 14.75 [±1.92] years), and

	• shortly before starting gender-affirming hormones (mean [±SD] age: 16.64 [±1.90] years). At follow up, the mean (±SD) CBCL scores were statistically significantly lower (improved) compared with baseline for: • Total T score (n=54, 60.70 [±12.76] versus 54.46 [±11.23], p<0.001 • Internalising T score (n=54, 61.00 [±12.21] versus 52.17 [±9.81], p<0.001) • Externalising T score (n=54, 58.04 [±12.99] versus 53.81 [±11.86], p=0.001). At follow up, the mean (±SD) YSR scores were statistically significantly lower (improved) compared with baseline for: • Total T score (n=54, 55.46 [±11.56] versus 50.00 [±10.56], p<0.001) • Internalising T score (n=54, 56.04 [±12.49] versus 49.78 [±11.63], p<0.001) • Externalising T score (n=54, 53.30 [±11.87] versus 49.98 [±9.35], p=0.009). The proportion of adolescents scoring in the clinical range decreased from baseline to follow up on the CBCL total problem scale (44.4% versus 22.2%, p=0.001) and the internalising scale of the YSR (29.6% versus 11.1%, p=0.017) **(VERY LOW)**. One study (Staphorsius et al. 2015) assessed CBCL in a cohort of adolescents with gender dysphoria (transfemale: n=18, mean [±SD] age 15.1 [±2.4] years and transmale: n=22, mean [±SD] age 15.8 [±1.9] years) either receiving GnRH analogues (transfemale, n=8 and transmale, n=12), or not receiving GnRH analogues (transfemale, n=10 and transmale, n=10). The mean (±SD) CBCL scores for each group were (statistical analysis unclear): • transfemales (total) 57.8 [±9.2] • transfemales receiving GnRH analogues 57.4 [±9.8] • transfemales not receiving GnRH analogues 58.2 [±9.3] • transmales (total) 60.4 [±10.2] • transmales receiving GnRH analogues 57.5 [±9.4] • transmales not receiving GnRH analogues 63.9 [±10.5] **(VERY LOW)**. **These studies provide very low certainty evidence that during treatment with GnRH analogues psychosocial functioning may improve, with the proportion of adolescents in the clinical range for some CBCL and YSR scores decreasing over time.**
Engagement with health care services **Certainty of evidence: very low**	This is an important outcome because patient engagement with health care services will impact on their clinical outcomes. Two uncontrolled observational cohort studies provided evidence relating to loss to follow up, which could be a marker of engagement with health care services (Brik et al. 2018 and Costa et al. 2015).

	In one retrospective study (Brik et al. 2018), 9 adolescents (9/214, 4.2%) who had stopped attending appointments were excluded from the study between November 2010 and July 2019 **(VERY LOW)**. One prospective study (Costa et al. 2015) had evidence for a large loss to follow-up over time. The sample size at baseline (T0) and 6 months (T1) was 201, which dropped by 39.8% to 121 after 12 months (T2) and by 64.7% to 71 at 18 months follow-up (T3). No explanation of the reasons for loss to follow-up are reported **(VERY LOW)**. Due to their design there was no reported loss to follow-up in the other 3 effectiveness studies (de Vries et al 2011; Khatchadourian et al. 2014; Staphorsius et al. 2015). **These studies provide very low certainty evidence about loss to follow up, which could be a marker of engagement with health care services, during treatment with GnRH analogues. Due to the large variation in rates between studies no conclusions could be drawn.**
Impact on extent of and satisfaction with surgery	This is an important outcome because some children and adolescents with gender dysphoria may proceed to transitioning surgery. No evidence was identified.

Stopping treatment
 	
Certainty of evidence: very low
	This is an important outcome because there is uncertainty about the short- and long-term safety and adverse effects of GnRH analogues in children and adolescents with gender dysphoria.
	
	Two uncontrolled, retrospective, observational cohort studies provided evidence relating to stopping GnRH analogues. One study had complete reporting of the cohort (Brik et al. 2018), the other (Khatchadourian et al. 2014) had incomplete reporting of its cohort, particularly for transfemales where outcomes for only 4/11 were reported.
	
	Brik et al. 2018 narratively reported the reasons for stopping GnRH analogues in a cohort of 143 adolescents (38 transfemales and 105 transmales). Median age at the start of GnRH analogues was 15.0 years (range, 11.1–18.6 years) in transfemales and 16.1 years (range, 10.1–17.9 years) in transmales. Of these adolescents, 125 (87%, 36 transfemales, 89 transmales) subsequently started gender-affirming hormones after 1.0 (0.5–3.8) and 0.8 (0.3–3.7) years of GnRH analogues. At the time of data collection, the median duration of GnRH analogue use was 2.1 years (1.6–2.8).
	
	During the follow-up period 6.3% (9/143) of adolescents had discontinued GnRH analogues after a median duration of 0.8 years (range 0.1 to 3.0). The percentages and reasons for stopping were:
	• 2.8% (4/143) stopped GnRH analogues although they wanted to continue endocrine treatments for gender dysphoria:
	  ○ 1 transmale stopped due to increase in mood problems, suicidal thoughts and confusion attributed to GnRH analogues
	  ○ 1 transmale had hot flushes, increased migraines, fear of injections, stress at school and unrelated medical issues, and temporarily stopped treatment (after 4 months) and restarted 5 months later.

	○ 1 transmale had mood swings 4 months after starting GnRH analogues. After 2.2 years had unexplained severe nausea and rapid weight loss and discontinued GnRH analogues after 2.4 years ○ 1 transmale stopped GnRH analogues because of inability to regularly collect medication and attend appointments for injections. • 3.5% (5/143) stopped treatment because they no longer wished to receive gender-affirming treatment for various reasons **(VERY LOW)**. Khatchadourian et al. 2014 narratively reported the reasons for stopping GnRH analogues in a cohort of 26 adolescents (15 transmales and 11 transfemales), 42% (11/26) discontinued GnRH analogues during follow-up between 1998 and 2011. Of 15 transmales receiving GnRH analogues, 14 received testosterone during the observation period, of which: • 7 continued GnRH analogues after starting testosterone • 7 stopped GnRH analogues after a median of 3.0 years (range 0.2 to 9.2 years), of which: ○ 5 stopped after hysterectomy and salpingo-oophorectomy ○ 1 stopped after 2.2 years (transitioned to gender-affirming hormones) ○ 1 stopped after <2 months due to mood and emotional lability **(VERY LOW)**. Of 11 transfemales receiving GnRH analogues, 5 received oestrogen during the observation period, of which: • 4 continued GnRH analogues after starting oestrogen • 1 stopped GnRH analogues when taking oestrogen (no reason reported) **(VERY LOW)**. Of the remaining 6 transfemales taking GnRH analogues: • 1 stopped GnRH analogues after a few months due to emotional lability • 1 stopped GnRH analogues before taking oestrogen (the following year delayed due to heavy smoking) • 1 stopped GnRH analogues after 13 months due not to pursuing transition **(VERY LOW)**. **These studies provide very low certainty evidence for the number of adolescents who stop GnRH analogues and the reasons for this.**

Abbreviations: GnRH, gonadotrophin releasing hormone; SD, standard deviation.

In children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement

Safety	
Change in bone density: lumbar **Certainty of evidence: very low**	This is an important outcome because puberty is an important time for bone development and puberty suppression may affect bone development, as shown by changes in lumbar bone density. Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on bone density (based on lumbar BMAD) between starting with a GnRH analogue and at 1 and 2 year intervals (Joseph et al. 2019), and between starting GnRH analogues and starting gender-affirming hormones (Klink et al. 2015 and Vlot et al. 2017). All outcomes were reported separately for transfemales and transmales; also see subgroups table below. BMAD is a size adjusted value of BMD incorporating body size measurements using UK norms in growing adolescents. It was reported as g/cm^3 and as z-scores. Z-scores report how many standard deviations from the mean a measurement sits. A z-score of 0 is equal to the mean, a z-score of −1 is equal to 1 standard deviation below the mean, and a z-score of +1 is equal to 1 standard deviation above the mean. One retrospective observational study (Joseph et al. 2019, n=70) provided non-comparative evidence on change in lumbar BMAD increase using z-scores. • The z-score for lumbar BMAD was statistically significantly lower at 2 years compared with baseline in transfemales (z-score [±SD]: baseline 0.486 [0.809], 2 years −0.279 [0.930], p=0.000) and transmales (baseline −0.361 [1.439], 2 years −0.913 [1.318], p=0.001) **(VERY LOW)**. • The z-score for lumbar BMAD was statistically significantly lower at 1 year compared with baseline in transfemales (baseline 0.859 [0.154], 1 year −0.228 [1.027], p=0.000) and transmales (baseline −0.186 [1.230], 1 year −0.541 [1.396], p=0.006) **(VERY LOW)**. • Actual lumbar BMAD values in g/cm^3 were not statistically significantly different between baseline and 1 or 2 years in transfemales or transmales **(VERY LOW)**. Two retrospective observational studies (Klink et al. 2015 and Vlot et al. 2017, n=104 in total) provided non-comparative evidence on change in lumbar BMAD between starting GnRH analogues and starting gender-affirming hormones. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. In Klink et al. 2015 the z-score for lumbar BMAD was not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales but was statistically significantly lower when starting gender-affirming hormones in transmales (z-score mean [±SD]: GnRH analogue 0.28 [±0.90], gender-affirming hormone −0.50 [±0.81], p=0.004). Actual lumbar BMAD values in g/cm^3 were not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales or transmales **(VERY LOW)**.

	Vlot et al. 2017 reported change from starting GnRH analogues to starting gender-affirming hormones in lumbar BMAD by bone age. • The z-score for lumbar BMAD in transfemales with a bone age of <15 years was statistically significantly lower at starting gender-affirming hormone treatment than at starting GnRH analogues (z-score median [range]: GnRH analogue −0.20 [−1.82 to 1.18], gender-affirming hormone −1.52 [−2.36 to 0.42], p=0.001) but was not statistically significantly different in transfemales with a bone age ≥15 years **(VERY LOW)**. • The z-score for lumbar BMAD in transmales with a bone age of <14 years was statistically significantly lower at starting gender-affirming hormone treatment than at starting GnRH analogues (z-score median [range]: GnRH analogue −0.05 [−0.78 to 2.94], gender-affirming hormone −0.84 [−2.20 to 0.87], p=0.003) and in transmales with a bone age ≥14 years (GnRH analogue 0.27 [−1.60 to 1.80], gender-affirming hormone −0.29 [−2.28 to 0.90], p≤0.0001) **(VERY LOW)**. • Actual lumbar BMAD values in g/cm^3 were not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales or transmales with young or old bone age **(VERY LOW)**. Two uncontrolled, observational, retrospective studies provided evidence for the effect of GnRH analogues on bone density (based on lumbar BMD) between starting GnRH analogues and either at 1 or 2 year intervals (Joseph et al. 2019), or starting gender-affirming hormones (Klink et al. 2015). All outcomes were reported separately for transfemales and transmales; also see subgroups table below. One retrospective observational study (Joseph et al. 2019, n=70) provided non-comparative evidence on change in lumbar BMD increase using z-scores. • The z-score for lumbar BMD was statistically significantly lower at 2 years compared with baseline in transfemales (z-score mean [±SD]: baseline 0.130 [0.972], 2 years −0.890 [±1.075], p=0.000) and transmales (baseline −0.715 [±1.406], 2 years −2.000 [1.384], p=0.000) **(VERY LOW)**. • The z-score for lumbar BMD was statistically significantly lower at 1 year compared with baseline in transfemales (z-score mean [±SD]: baseline −0.016 [±1.106], 1 year −0.461 [±1.121], p=0.003) and transmales (baseline −0.395 [±1.428], 1 year −1.276 [±1.410], p=0.000) **(VERY LOW)**. • With the exception of transmales, where lumbar BMD in kg/m^2 increased between baseline and 1 year (mean [±SD]: baseline 0.694 [±0.149], 1 year 0.718 [±0.124], p=0.006), actual lumbar BMD values were not statistically significantly different between baseline and 1 or 2 years in transfemales or between 0 and 2 years in transmales **(VERY LOW)**. One retrospective observational study (Klink et al. 2015, n=34) provided non-comparative evidence on change in lumbar BMD between starting GnRH analogues and starting gender-affirming hormones. • The z-score for lumbar BMD was not statistically significantly different between starting GnRH analogue and starting gender-affirming hormone treatment in transfemales, but was

	statistically significantly lower when starting gender-affirming hormones in transmales (z-score mean [±SD]: GnRH analogue 0.17 [±1.18], gender-affirming hormone −0.72 [±0.99], p<0.001) **(VERY LOW)**. • Actual lumbar BMD in g/cm^2 was not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales but was statistically significantly lower when starting gender-affirming hormones in transmales (mean [±SD]: GnRH analogues 0.95 [±0.12], gender-affirming hormones 0.91 [±0.10], p=0.006) **(VERY LOW)**. **These studies provide very low certainty evidence that GnRH analogues reduce the expected increase in lumbar bone density (BMAD or BMD) compared with baseline (although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically significantly decrease actual lumbar bone density (BMAD or BMD).**
Change in bone density: femoral **Certainty of evidence: very low**	This is an important outcome because puberty is an important time for bone development and puberty suppression may affect bone development, as shown by changes in femoral bone density. Two uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on bone density (based on femoral BMAD) between starting treatment with a GnRH analogue and starting gender-affirming hormones (Klink et al. 2015 and Vlot et al. 2017). All outcomes were reported separately for transfemales and transmales; also see subgroups table below. One retrospective observational study (Klink et al. 2015, n=34) provided non-comparative evidence on change in femoral area BMAD between starting GnRH analogues and starting gender-affirming hormones. All outcomes were reported separately for transfemales and transmales. • The z-score for femoral area BMAD was not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales or transmales **(VERY LOW)**. • Actual femoral area BMAD values were not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transmales or transfemales **(VERY LOW)**. One retrospective observational study (Vlot et al. 2017, n=70) provided non-comparative evidence on change in femoral neck (hip) BMAD between starting GnRH analogues and starting gender-affirming hormones. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. • The z-score for femoral neck BMAD in transfemales with a bone age of <15 years was not statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (z-score median [range]: GnRH analogue −0.71 [−3.35 to 0.37], gender-affirming hormone −1.32 [−3.39 to 0.21], p≤0.1) or in transfemales with a bone age ≥15 years (GnRH analogue −0.44 [−1.37 to 0.93], gender-affirming hormone −0.36 [−1.50 to 0.46]) **(VERY LOW)**.

	• The z-score for femoral neck BMAD in transmales with a bone age of <14 years was not statistically significantly lower at starting gender-affirming hormones than at starting GnRH analogues (z-score median [range]: GnRH analogue −0.01 [−1.30 to 0.91], gender-affirming hormone −0.37 [−2.28 to 0.47]) but was statistically significantly lower in transmales with a bone age ≥14 years (GnRH analogue 0.27 [−1.39 to 1.32], gender-affirming hormone −0.27 [−1.91 to 1.29], p=0.002) **(VERY LOW)**. • Actual femoral neck BMAD values were not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales or in transmales with a young bone age, but were statistically significantly lower in transmales with a bone age ≥14 years (GnRH analogue 0.33 [0.25 to 0.39), gender-affirming hormone 0.30 [0.23 to 0.41], p≤0.01) **(VERY LOW)**. Two uncontrolled, observational, retrospective studies provided evidence for the effect of GnRH analogues on bone density (based on femoral BMD) between starting GnRH analogues and either at 1 or 2 year intervals (Joseph et al. 2019), or starting gender-affirming hormones (Klink et al. 2015). All outcomes were reported separately for transfemales and transmales; also see subgroups table below. One retrospective observational study (Joseph et al. 2019, n=70) provided non-comparative evidence on change in femoral neck BMD increase using z-scores. All outcomes were reported separately for transfemales and transmales. • The z-score for femoral neck BMD was statistically significantly lower at 2 years compared with baseline in transfemales (z-score mean [±SD]: baseline 0.0450 [±0.781], 2 years −0.600 [±1.059], p=0.002) and transmales (baseline −1.075 [±1.145], 2 years −1.779 [±0.816], p=0.001) **(VERY LOW)**. • The z-score for femoral neck BMD was statistically significantly lower at 1 year compared with baseline in transfemales (z-score mean [±SD]: baseline 0.157 [±0.905], 1 year −0.340 [±0.816], p=0.002) and transmales (baseline −0.863 [±1.215], 1 year −1.440 [±1.075], p=0.000) **(VERY LOW)**. • Actual femoral neck BMD values in kg/m^2 were not statistically significantly different between baseline and 1 or 2 years in transmales or transfemales **(VERY LOW)**. One retrospective observational study (Klink et al. 2015, n=34) provided non-comparative evidence on change in femoral area BMD between starting GnRH analogues and starting gender-affirming hormones. All outcomes were reported separately for transfemales and transmales. • The z-score for femoral area BMD was not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales, but was statistically significantly lower in transmales (z-score mean [±SD]: GnRH analogue 0.36 [±0.88], gender-affirming hormone −0.35 [±0.79], p=0.001) **(VERY LOW)**. • Actual femoral area BMD values were not statistically significantly different between starting GnRH analogues and starting gender-affirming hormones in transfemales, but were

	statistically significantly lower in transmales (mean [±SD] GnRH analogue 0.92 [±0.10], gender-affirming hormone 0.88 [±0.09], p=0.005) **(VERY LOW)**. **These studies provide very low certainty evidence that GnRH analogues may reduce the expected increase in femoral bone density (femoral neck or area BMAD or BMD) compared with baseline (although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically significantly decrease actual femoral bone density (femoral area BMAD or femoral neck BMD), apart from actual femoral area BMD in transmales.**
Cognitive development or functioning **Certainty of evidence: very low**	This is an important outcome because puberty is an important time for cognitive development and puberty suppression may affect cognitive development or functioning. One cross-sectional observational study (Staphorsius et al. 2015, n=70) provided comparative evidence on cognitive development or functioning in adolescents with gender dysphoria on GnRH analogues compared with adolescents with gender dysphoria not on GnRH analogues. Cognitive functioning was measured using an IQ test. Reaction time (in seconds) and accuracy (percentage of correct trials) were measured using the Tower of London (ToL) task. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. No statistical analyses or interpretation of the results in these groups were reported: • IQ in transfemales (mean [±SD] GnRH analogue 94.0 [±10.3], control 109.4 [±21.2]). IQ transmales (GnRH analogue 95.8 [±15.6], control 98.5 [±15.9]. • Reaction time in transfemales (mean [±SD] GnRH analogue 10.9 [±4.1], control: 9.9 [±3.1]). Reaction time transmales (GnRH analogue 9.9 [±3.1], control 10.0 [±2.0]). • Accuracy score in transfemales (GnRH analogue 73.9 [±9.1], control 83.4 [±9.5]. Accuracy score in transmales (GnRH analogue 85.7 [±10.5], control 88.8 [±9.7]. **This study provides very low certainty evidence (with no statistical analysis) on the effects of GnRH analogues on cognitive development or functioning. No conclusions could be drawn.**
Other safety outcomes: kidney function **Certainty of evidence: very low**	This is an important outcome because if renal damage (raised serum creatinine is a marker of this) is suspected, GnRH analogues may need to be stopped. One prospective observational study (Schagen et al. 2016, n=116) provided non-comparative evidence on change in serum creatinine between starting GnRH analogues and at 1 year. All outcomes were reported separately for transfemales and transmales; also see subgroups table below. • There was no statistically significant difference between baseline and 1 year for serum creatinine in transfemales (mean [±SD] baseline 70 [±12], 1 year 66 [±13], p=0.20). • There was a statistically significant decrease between baseline and 1 year for serum creatinine in transmales (baseline 73 [±8], 1 year 68 [±13], p=0.01).

	This study provides very low certainty evidence that GnRH analogues do not affect renal function.
Other safety outcomes: liver function **Certainty of evidence: very low**	This is an important outcome because if treatment-induced liver injury (raised liver enzymes are a marker of this) is suspected, GnRH analogues may need to be stopped. One prospective observational study (Schagen et al. 2016, n=116) provided non-comparative evidence on elevated liver enzymes between starting GnRH analogues and during use. No comparative values or statistical analyses were reported. • Glutamyl transferase was not elevated at baseline or during use in any person. • Mild elevations of AST and ALT above the reference range were present at baseline but were not more prevalent during use than at baseline. • Glutamyl transferase, AST, and ALT levels did not significantly change from baseline to 12 months of use. **This study provides very low certainty evidence (with no statistical analysis) that GnRH analogues do not affect liver function.**
Other safety outcomes: adverse effects **Certainty of evidence: very low**	This is an important outcome because if there are adverse effects, GnRH analogues may need to be stopped. One uncontrolled, retrospective, observational cohort study (Khatchadourian et al. 2014) provided evidence relating to adverse effects from GnRH analogues. It had incomplete reporting of its cohort, particularly for transfemales where outcomes for only 4/11 were reported. Khatchadourian et al. 2014 reported adverse effects in a cohort of 26 adolescents (15 transmales and 11 transfemales) receiving GnRH analogues. Of these: • 1 transmale developed sterile abscesses; they were switched from leuprolide acetate to triptorelin, and this was well tolerated. • 1 transmale developed leg pains and headaches, which eventually resolved • 1 participant gained 19 kg within 9 months of starting GnRH analogues. **This study provides very low certainty evidence about potential adverse effects of GnRH analogues. No conclusions could be drawn.**

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; BMAD, bone mineral apparent density; BMD, bone mineral density; GnRH, gonadotrophin releasing hormone; IQ, intelligence quotient; NS, not significant; SD, standard deviation.

In children and adolescents with gender dysphoria, what is the cost-effectiveness of GnRH analogues compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?

Outcome	Evidence statement

Cost-effectiveness	No studies were identified to assess the cost-effectiveness of GnRH analogues for children and adolescents with gender dysphoria.

From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may benefit from GnRH analogues more than the wider population of interest?

Subgroup	Evidence statement
Sex assigned at birth males (transfemales) **Certainty of evidence: Very low**	Some studies reported data separately for sex assigned at birth males (transfemales). This included some direct comparisons with sex assigned at birth females (transmales). **Impact on gender dysphoria** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence for gender dysphoria in sex assigned at birth males. See the clinical effectiveness results table above for a full description of the study. The mean (±SD) UGDS score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean UGDS score [±SD]: 47.95 [±9.70] versus 56.57 [±3.89]) and T1 (n=not reported, 49.67 [±9.47] versus 56.62 [±4.00]); between sex difference p<0.001 **(VERY LOW)**. One further prospective observational longitudinal study (Costa et al. 2015) provided evidence for the impact on gender dysphoria in sex assigned at birth males. See the clinical effectiveness results table above for a full description of the study. Sex assigned at birth males had a statistically significantly lower (improved) mean (±SD) UGDS score of 51.6 [±9.7] compared with sex assigned at birth females (56.1 [±4.3], p<0.001). However, it was not reported if this was baseline or follow-up **(VERY LOW)**. **These studies provide very low certainty evidence that in sex assigned at birth males (transfemales), gender dysphoria is lower than in sex assigned at birth females (transmales).** **Impact on mental health** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence for the impact on mental health (depression, anger and anxiety) in sex assigned at birth males. See the clinical effectiveness results table above for a full description of the study. • The mean (±SD) depression (BDI-II) score was not statistically significantly different in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean BDI score [±SD]: 5.71 [±4.31] versus 10.34 [±8.24]) and T1 (n=not reported, 3.50 [±4.58] versus 6.09 [±7.93]), between sex difference p=0.057 • The mean (±SD) anger (TPI) score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean TPI score [±SD]: 5.22 [±2.76]

versus 6.43 [±2.78]) and T1 (n=not reported, 5.00 [±3.07] versus 6.39 [±2.59]), between sex difference p=0.022
- The mean (±SD) anxiety (STAI) score was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean STAI score [±SD]: 4.33 [±2.68] versus 7.00 [±2.36]) and T1 (n=not reported, 4.39 [±2.64] versus 6.17 [±2.69]), between sex difference p<0.001 **(VERY LOW)**.

This study provides very low certainty evidence that the impact on mental health (depression, anger and anxiety) may be different in sex assigned at birth males (transfemales) compared with sex assigned at birth females (transmales). Over time there was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for depression. However, sex assigned at birth males had statistically significantly lower levels of anger and anxiety than sex assigned at birth females at both baseline and follow up.

Impact on body image
One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on body image in sex assigned at birth males.
- The mean (±SD) BIS score for primary sex characteristics was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean BIS score [±SD]: 4.02 [±0.61] versus 4.16 [±0.52]) and T1 (n=not reported, 3.74 [±0.78] versus 4.17 [±0.58]), between sex difference p=0.047
- The mean (±SD) BIS score for secondary sex was statistically significantly lower (improved) in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean BIS score [±SD]: 2.66 [±0.50] versus 2.81 [±0.76]) and T1 (n=not reported, 2.39 [±0.69] versus 3.18 [±0.42]), between sex difference p=0.001
- The mean (±SD) BIS score for neutral body characteristics was not statistically significantly different in sex assigned at birth males compared with sex assigned at birth females at both baseline (T0) (n=not reported, mean BIS score [±SD]: 2.60 [±0.58] versus 2.24 [±0.62]) and T1 (n=not reported, 2.32 [±0.59] versus 2.61 [±0.50]), between sex difference p=0.777 **(VERY LOW)**.

This study provides very low certainty evidence that the impact on body image may be different in sex assigned at birth males (transfemales) compared with sex assigned at birth females (transmales). Sex assigned at birth males are less dissatisfied with their primary and secondary sex characteristics than sex assigned at birth females at both baseline and follow up, but the satisfaction with neutral body characteristics is not different.

Psychosocial impact
One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence for psychosocial impact in terms

	of global functioning (CGAS) and psychosocial functioning (CBCL and YSR) in sex assigned at birth males. • Sex assigned at birth males had statistically higher mean (±SD) CGAS scores compared with sex assigned at birth females at both baseline (T0) (n=54, 73.10 [±8.44] versus 67.25 [±11.06]) and T1 (n=54, 77.33 [±8.69] versus 70.30 [±9.44]), between sex difference p=0.021 • There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for the CBCL Total T score at T0 or T1 (n=54, p=0.110) • There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for the CBCL internalising T score at T0 or T1 (n=54, p=0.286) • Sex assigned at birth males had statistically lower mean (±SD) CBCL externalising T scores compared with sex assigned at birth females at both T0 (n=54, 54.71 [±12.91] versus 60.70 [±12.64]) and T1 (n=54, 48.75 [±10.22] versus 57.87 [±11.66]), between sex difference p=0.015 • There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for the YSR Total T score at T0 or T1 (n=54, p=0.164) • There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for the YSR internalising T score at T0 or T1 (n=54, p=0.825) • Sex assigned at birth males had statistically lower mean (±SD) YSR externalising T scores compared with sex assigned at birth females at both T0 (n=54, 48.72 [±11.38] versus 57.24 [±10.59]) and T1 (n=54, 46.52 [±9.23] versus 52.97 [±8.51]), between sex difference p=0.004 **(VERY LOW)**. One uncontrolled, observational, prospective cohort study (Costa et al. 2015) provided evidence for psychosocial impact in terms of global functioning (CGAS) in sex assigned at birth males. • Sex assigned at birth males had statistically significant lower mean (±SD CGAS scores at baseline) compared with sex assigned at birth females (n=201, 55.4 [±12.7] versus 59.2 [±11.8], p=0.03) **(VERY LOW)**. **These studies provide very low certainty evidence that psychosocial impact may be different in sex assigned at birth males (transfemales) compared with sex assigned at birth females (transmales). However, no conclusions could be drawn.** **Change in bone density: lumbar** Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on lumbar bone density in sex assigned at birth males (Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results. **These studies provide very low certainty evidence that GnRH analogues reduce the expected increase in lumbar bone density (BMAD or BMD) in sex assigned at birth males (transfemales; although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically**

	significantly decrease actual lumbar bone density (BMAD or BMD) in sex assigned at birth males (transfemales). **Change in bone density: femoral** Three uncontrolled, observational, retrospective studies provided evidence for the effect of GnRH analogues on femoral bone density in sex assigned at birth males (Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results. **These studies provide very low certainty evidence that GnRH analogues may reduce the expected increase in femoral bone density (femoral neck or area BMAD or BMD) in sex assigned at birth males (transfemales; although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically significantly decrease actual femoral bone density (femoral area BMAD or femoral neck BMD) in sex assigned at birth males (transfemales).** **Cognitive development or functioning** One cross-sectional observational study (Staphorsius et al. 2015) provided comparative evidence on cognitive development or functioning in sex assigned at birth males. See the safety results table above for a full description of the results. **This study provides very low certainty evidence (with no statistical analysis) on the effects of GnRH analogues on cognitive development or functioning in sex assigned at birth males (transfemales). No conclusions could be drawn.** **Other safety outcomes: kidney function** One prospective observational study (Schagen et al. 2016) provided non-comparative evidence on change in serum creatinine in sex assigned at birth males. See the safety results table above for a full description of the results. **This study provides very low certainty evidence that GnRH analogues do not affect renal function in sex assigned at birth males (transfemales).**
Sex assigned at birth females (transmales) **Certainty of evidence: Very low**	Some studies reported data separately for sex assigned at birth females (transmales). This included some direct comparisons with sex assigned at birth males (transfemales). **Impact on gender dysphoria** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) and one prospective observational longitudinal study (Costa et al. 2015) provided evidence for gender dysphoria in sex assigned at birth females. See the sex assigned at birth males (transfemales) row above for a full description of the results. **These studies provide very low certainty evidence that in sex assigned at birth females (transmales), gender dysphoria is higher than in sex assigned at birth males (transfemales) at both baseline and follow up.**

	Impact on mental health One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on mental health (depression, anger and anxiety) in sex assigned at birth females. See the sex assigned at birth males (transfemales) row above for a full description of the results. **This study provides very low certainty evidence that the impact on mental health (depression, anger and anxiety) may be different in sex assigned at birth females (transmales) compared with sex assigned at birth males (transfemales). Over time there was no statistically significant difference between sex assigned at birth females and sex assigned at birth males for depression. However, sex assigned at birth females had statistically significantly greater levels of anger and anxiety than sex assigned at birth males at baseline and follow up.** **Impact on body image** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence relating to the impact on body image in sex assigned at birth females. See the sex assigned at birth males (transfemales) row above for a full description of the results. **This study provides very low certainty evidence that the impact on body image may be different in sex assigned at birth females (transmales) compared with sex assigned at birth males (transfemales). Sex assigned at birth females are more dissatisfied with their primary and secondary sex characteristics than sex assigned at birth males at both baseline and follow up, but the satisfaction with neutral body characteristics is not different.** **Psychosocial impact** One uncontrolled prospective observational longitudinal study (de Vries et al. 2011) provided evidence for psychosocial impact in terms of global functioning (CGAS) and psychosocial functioning (CBCL and YSR) in sex assigned at birth females. One uncontrolled, observational, prospective cohort study (Costa et al. 2015) provided evidence for psychosocial impact in terms of global functioning (CGAS) in sex assigned at birth females. See the sex assigned at birth males (transfemales) row above for a full description of the results. **These studies provide very low certainty evidence that psychosocial impact may be different in sex assigned at birth females (transmales) compared with sex assigned at birth males (transfemales). However, no conclusions could be drawn.** **Change in bone density: lumbar** Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on lumbar bone density in sex assigned at birth females (Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results.

	These studies provide very low certainty evidence that GnRH analogues reduce the expected increase in lumbar bone density (BMAD or BMD) in sex assigned at birth females (transmales; although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically significantly decrease actual lumbar bone density (BMAD or BMD) in sex assigned at birth females (transmales). **Change in bone density: femoral** Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on femoral bone density in sex assigned at birth females (Joseph et al. 2019, Klink et al. 2015 and Vlot et al. 2017). See the safety results table above for a full description of the results. **These studies provide very low certainty evidence that GnRH analogues may reduce the expected increase in femoral bone density (femoral neck or area BMAD or BMD) in sex assigned at birth females (transmales; although some findings were not statistically significant). These studies also show that GnRH analogues do not statistically significantly decrease actual femoral bone density (femoral area BMAD or femoral neck BMD) in sex assigned at birth females (transmales), apart from actual femoral area.** **Cognitive development or functioning** One cross-sectional observational study (Staphorsius et al. 2015) provided comparative evidence on cognitive development or functioning in sex assigned at birth females. See the safety results table above for a full description of the results. **This study provides very low certainty evidence (with no statistical analysis) on the effects of GnRH analogues on cognitive development or functioning in sex assigned at birth females (transmales). No conclusions could be drawn.** **Other safety outcomes: kidney function** One prospective observational study (Schagen et al. 2016) provided non-comparative evidence on change in serum creatinine in sex assigned at birth females (transmales). See the safety results table above for a full description of the results. **This study provides very low certainty evidence that GnRH analogues do not affect renal function in sex assigned at birth females (transmales).**
Duration of gender dysphoria	No evidence was identified.
Age at onset of gender dysphoria	No evidence was identified.
Age at which GnRH analogue started	No evidence was identified.
Age at onset of puberty	No evidence was identified.

Tanner stage at which GnRH analogue started	No evidence was identified.
Diagnosis of autistic spectrum disorder	No evidence was identified.
Diagnosis of mental health condition	No evidence was identified.

Abbreviations: BDI-II, Beck Depression Inventory-II; BIS, Body Image Scale; CBCL, Child Behaviour Checklist; CGAS, Children's Global Assessment Scale; SD, standard deviation; STAI, Trait Anxiety Scale of the State-Trait Personality Inventory; TPI, Trait Anger Scale of the State-Trait Personality Inventory; UGDS, Utrecht Gender Dysphoria Scale; YSR, Youth Self-Report

From the evidence selected,

(a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?

(b) what were the ages at which participants commenced treatment with GnRH analogues?

(c) what was the duration of treatment with GnRH analogues?

Outcome	Evidence statement
Diagnostic criteria	In 5 studies (Costa et al. 2015, Klink et al. 2015, Schagen et al. 2016, Staphorsius et al. 2015 and Vlot et al. 2017) the DSM-IV-TR criteria of gender identity disorder was used. The study by Brik et al. 2020 used DSM-V criteria. The DSM-V has one overarching definition of gender dysphoria with separate specific criteria for children and for adolescents and adults. The general definition describes a conflict associated with significant distress and/or problems functioning associated with this conflict between the way they feel and the way they think of themselves which must have lasted at least 6 months. It was not reported how gender dysphoria was defined in the remaining 3 studies **(VERY LOW)**. **From the evidence selected, all studies that reported diagnostic criteria for gender dysphoria (6/9 studies) used the DSM criteria in use at the time the study was conducted.**
Age when GnRH analogues started	8/9 studies reported the age at which participants started GnRH analogues, either as the mean age (with SD) or median age (with the range): **Study** / **Mean age (±SD)** Costa et al. 2015 — 16.5 years (±1.3) de Vries et al. 2011 — 13.6 years (±1.8) Joseph et al. 2019 — 13.2 years (±1.4) in transfemales; 12.6 years (±1.0) in transmales Khatchadourian et al. 2014 — 14.7 years (±1.9)

<table>
<tr><td></td><td>
<table>
<tr><td>Klink et al. 2015</td><td>14.9 years (±1.9) in transfemales
15.0 years (±2.0) in transmales</td></tr>
</table>
<table>
<tr><th>Study</th><th>Median age (range)</th></tr>
<tr><td>Brik et al. 2020</td><td>15.5 years (11.1–18.6) in transfemales
16.1 years (10.1–17.9) in transmales</td></tr>
<tr><td>Schagen et al. 2016</td><td>13.6 years (11.6–17.9) in transfemales
14.2 years (11.1–18.6) in transmales</td></tr>
<tr><td>Vlot et al. 2017</td><td>13.5 years (11.5–18.3) in transfemales
15.1 years (11.7–18.6) in transmales</td></tr>
</table>
Age at the start of GnRH analogues was not reported in Staphorsius et al. 2015, but participants were required to be at least 12 years (VERY LOW).

The evidence included showed wide variation in the age (11 to 18 years old) at which children and adolescents with gender dysphoria started GnRH analogues.
</td></tr>
<tr><td>Duration of treatment</td><td>The duration of treatment with GnRH analogues was reported in 3/9 studies. The median duration was:
• 2.1 years (range 1.6–2.8) in Brik et al. 2020.
• 1.3 years (range 0.5–3.8) in transfemales and 1.5 years (range 0.25–5.2) in transmales in Klink et al. 2015.

In Staphorsius et al. 2015, the mean duration was 1.6 years (SD ±1.0).

In de Vries et al. 2011, the mean duration of time between starting GnRH analogues and gender-affirming hormones was 1.88 years (SD ±1.05).

The evidence included showed wide variation in the duration of treatment with GnRH analogues, but most studies did not report this information. Treatment duration ranged from a few months up to about 5 years.</td></tr>
</table>

Abbreviations: DSM, Diagnostic and Statistical Manual of Mental Disorders criteria; SD, standard deviation.

6. Discussion

A key limitation to identifying the effectiveness and safety of GnRH analogues for children and adolescents with gender dysphoria is the lack of reliable comparative studies. The lack of clear, expected outcomes from treatment with a GnRH analogue (the purpose of which is to suppress secondary sexual characteristics which may cause distress from unwanted pubertal changes) also makes interpreting the evidence difficult. The size of the population with gender dysphoria means conducting a prospective trial may be unrealistic, at least on a single centre basis. There may also be ethical issues with a 'no treatment arm' in comparative trials of GnRH analogues, where there may be poor mental health outcomes if treatment is withheld. However, the use of an active comparator such as close psychological support may reduce ethical concerns in future trials.

The studies included in this evidence review are all small, uncontrolled observational studies, which are subject to bias and confounding, and are of very low certainty as

assessed using modified GRADE. All the included studies reported physical and mental health comorbidities and concomitant treatments very poorly. For example, very little data are reported on how many children and adolescents needed additional mental health support, and for what reasons, or whether additional interventions, and what form and duration (for example drug treatment or counselling) that took. This is a possible confounder for the treatment outcomes in the studies because changes in critical and important outcomes may be attributable to external care rather than the psychological support or GnRH analogues used in the studies.

The studies that reported diagnostic criteria for gender dysphoria (6/9 studies) used the Diagnostic and Statistical Manual of Mental Disorders (DSM) criteria in use at the time the study was conducted (either DSM-IV-TR or DSM-V). The definition was unclear in the remaining studies. There was wide variation in the ages at which participants started a GnRH analogue, typically ranging from about 11 to 18 years. Similarly, there was a wide variation in the duration of use, but few studies reported this.

Changes in outcome scores for clinical effectiveness were assessed for statistical significance in the 3 studies reporting these outcomes (Costa et al. 2015; de Vries et al. 2011; Staphorsius et al. 2015). However, there is relatively little interpretation of whether the changes in outcome scores seen in these studies are clinically meaningful.

For some outcomes there was no statistically significant difference from before starting GnRH analogues until just before starting gender-affirming hormones. These were the Utrecht Gender Dysphoria Scale (UGDS) (which was assessed in 1 study de Vries et al. 2011), the Trait Anger (TPI) and Trait Anxiety (STAI) Scales (which were assessed in 1 study de Vries et al. 2011), and Body Image Scale (BIS) which was assessed in 1 study (de Vries et al. 2011).

The Beck Depression Inventory (BDI-II) was used in 1 study (de Vries et al. 2011) to assess change in depression from before starting GnRH analogues to just before starting gender-affirming hormones. The result is statistically significant, with the mean (±SD) BDI-II score decreasing from 8.31 (±7.12) at baseline to 4.95 (±6.27) at follow up (p=0.004). However, both scores fall into the minimal range using the general guidelines for interpretation of BDI-II (0 to 13 minimal, 14 to 19 mild depression, 20 to 28 moderate depression and 29 to 63 severe depression), suggesting that while statistically significant, it is unclear if this is a clinically meaningful change.

Psychosocial outcomes were assessed in 3 studies (Costa et al. 2015; de Vries et al. 2011; Staphorsius et al. 2015) using the Children's Global Assessment Scale (CGAS) and Child Behavior Checklist/Youth Self-Report (CBCL/YSR). The CGAS score was assessed in 2 studies (Costa et al. 2015; de Vries et al. 2011). In de Vries et al. 2011 the mean (±SD) CGAS score statistically significantly increased over time from 70.24 [±10.12] at baseline to 73.90 [±9.63] at follow up. CGAS scores are clinically categorised into 10 categories (10 to 1, 20 to 11 and so on until 100 to 91) and both scores reported were in a single category (71 to 80, no more than slight impairment) suggesting that while statistically significant, it is unclear if this is a clinically meaningful change. The Costa et al. 2015 study does highlight a larger change in CGAS scores from baseline to follow-up (mean [±SD] 58.72 [±11.38] compared with 67.40 [±13.39]), but whether this is clinically meaningful is unclear. The average score moved from the clinical category of 60 to 51 (variable functioning with sporadic difficulties) at baseline to 70 to 61 (some difficulty in a single area, but generally

functioning pretty well) at follow up, but the large standard deviations suggest clinically significant overlaps between the scores from baseline to follow-up.

Psychosocial functioning using the CBCL/YSR was assessed in 2 studies (de Vries et al. 2011; Staphorsius et al. 2015). In de Vries et al. 2011 there was a statistically significant reduction in both CBCL and YSR scores from before starting GnRH analogues to just before starting gender-affirming hormones. The study interpreted the CBCL/YSR with a proportion of adolescents who scored in the clinical range (a T-score above 63), which allows changes in clinically meaningful scores to be assessed, and proportions of adolescents in the clinical range for some CBCL and YSR scores decreased over time. One cross-sectional study (Staphorsius et al. 2015) assessed CBCL scores only, but it was unclear if this was the Total T score, or whether subscales of internalising or externalising scores were also assessed, and whether the results were statistically significant.

The 2 prospective observational studies (Costa et al. 2015; de Vries et al. 2011) are confounded by a number of common factors. Firstly, the single assessment of scores at baseline means it is unclear if scores were stable, already improving or declining before starting treatment. Secondly, in an uncontrolled study any changes in scores from baseline to follow-up could be attributed to a regression-to-mean, for example getting older has been positively associated with maturity and wellbeing. The studies use mean and standard deviations in the descriptive statistics and analyses; however, they do not report testing the normality of data which would support the use of parametric measures. The study by de Vries et al. 2011 used general linear models (regression) to examine between and within group variances (changes in outcomes). In using such models, the data is assumed to be balanced (measured at regular intervals and without missing data), but the large ranges in ages at which participants were assessed and started on various interventions suggests that ascertainment of outcome was unlikely to be regular and missing data was likely. Missing data was handled through listwise deletion (omits those cases with the missing data and analyses the remaining data) which is acceptable if data loss is completely random but for some outcomes where there was incomplete data for individual items this was not random (items were introduced by the authors after the first eligible adolescents had started GnRH analogues). The study provided no detail on whether these assumptions for the modeling were met, they also provided no adequate assessment of whether any regression diagnostics (analysis that seek to assess the validity of a model) or model fit (how much of the variance in outcome is explained by the between and within group variance) were undertaken.

The 2 retrospective observational studies (Brik et al. 2020; Khatchadourian et al. 2014) both only report absolute numbers for each trajectory along with reasons for stopping GnRH analogues. It is difficult to assess outcomes from such single centre studies because there is little comparative data for outcomes from other such services. A lack of any critical or other important outcomes also means the success of the treatment across all the participants is difficult to judge.

Three uncontrolled, observational, retrospective studies provided evidence relating to the effect of GnRH analogues on bone density (Joseph et al. 2019; Klink et al. 2015; Vlot et al. 2017). In all 3 studies, the participants acted as their own controls and change in bone density was determined between starting GnRH analogues and either after 1 and 2 year follow-up timepoints (Joseph et al. 2019) or when gender-affirming hormones were started

(Klink et al. 2015 and Vlot et al. 2017). Observational studies such as these can only show an association with GnRH analogues and bone density; they cannot show that GnRH analogues caused any differences in bone density seen. Because there was no comparator group and participants acted as their own controls, it is unclear whether the findings are associated with GnRH analogues or due to changes over time. The authors reported z-scores which allows for comparison with the expected increase in bone density in the general population. However, because no concomitant treatments or comorbidities were reported it is possible that the findings may not be because of GnRH analogues and there is another way in which the study population differs from the general population.

All the studies are from a limited number of, mainly European, care facilities. They are described as either tertiary referral or expert services but the low number of services providing such care and publishing evidence may bias the results towards the outcomes in these services only and limit extrapolation.

The first study (Brik et al. 2020) was an uncontrolled, retrospective, observational study that assessed the outcome trajectories of adolescents receiving GnRH analogues for gender dysphoria. This study followed-up 143 individuals who had received GnRH analogues (38 transfemales and 105 transmales) using clinical records to show outcomes for up to 9 years (continuing use of GnRH analogues, reasons for stopping GnRH analogues and onward care such as gender-affirming hormone use). The methods and results are well reported, but no analysis of data was undertaken. The views of adolescents and their parents are particularly difficult to interpret because no data on how many responded to each question and in what ways are reported.

The second study (Costa et al. 2015) was an uncontrolled, prospective observational study which assessed global functioning in adolescents with gender dysphoria using CGAS every 6 months, including during the first 6 months where statistically significant improvements were seen without GnRH analogues. The study is confounded by significant unexplained loss to follow-up (64.7%: from n=201 adolescents to n=71 after 18 months). Missing data for those lost to follow-up maybe more than sufficient to change the direction of effects seen in the study if the reasons for loss to follow-up are systematic (such as deriving little or no benefit from treatment). The study uses clustered data in its analysis, a single outcome (CGAS) measured in clusters (at different visits), and the analysis does not take account of the correlation of scores (data at different time points are not independent) as a significant change in scores early in the study means the successive changes measured against baseline were also significant. The study relies on multiple (>20) pairwise independent *t*-tests to examine change in CGAS between the 4 time points, increasing the possibility of type-I error (a false positive which occurs when a researcher incorrectly rejects a true null hypothesis) because the more tests performed the more likely a statistically significant result will be observed by chance alone.

The Costa et al. 2015 study compares immediately eligible and delayed eligible cohorts, however, it is highly likely that they are non-comparable groups because the immediately eligible group were those able to start GnRH analogues straight away whilst those in the delayed eligible group were either not ready to make a decision about starting treatment (no age comparison was made between the 2 groups so it is unclear if they were a younger cohort than the immediately eligible group) or had comorbid mental health or psychological difficulties. The authors report that those with concomitant problems (such as mental health

problems, substantial problems with peers, or conflicts with parents or siblings) were referred to local mental health services but no details are provided.

The third study (de Vries et al. 2011) was an uncontrolled, prospective observational study which assessed gender dysphoria and psychological functioning before and after puberty suppression in adolescents with gender dysphoria. Although the study mentions the DSM-IV-TR there is no explicit discussion of this, or any other criteria, being used as the diagnostic criteria for study entry. There are no details reported for how the outcomes in the study were assessed, and by whom. The length of follow-up for the outcomes in the model are questionable in relation to whether there was sufficient time for GnRH analogues to have a measurable effect. The time points used are start of GnRH analogues and start of gender-affirming hormones. Overall, the mean time between starting GnRH analogues and gender-affirming hormones was 1.88 (±1.05) years, but the range is as low as just 5 months between the 2 time points, which may be insufficient for any difference in outcome to have occurred in some individuals.

The fourth study (Joseph et al. 2019) was a retrospective, longitudinal observational single centre study which assessed bone mineral density in adolescents with gender dysphoria in the UK. For inclusion in the study, participants had to have been assessed by the Gender Identity Development Service multi-disciplinary psychosocial health team for at least 4 assessments over a minimum of 6 months. No other diagnostic criteria, such as the DSM-IV-TR, are discussed. Bone density was assessed using dual energy X-ray absorptiometry (DAXA) scan of the lumbar spine (L1-L4) and the femoral neck at baseline (n=70), 1 year (n=70) and 2 years after starting GnRH analogues (n=39). The results suggest a possible association between GnRH analogues and bone mineral apparent density. However, the evidence is of poor quality, and the results could be due to bias or chance. No concomitant treatments or comorbidities were reported.

The fifth study (Khatchadourian et al. 2014) was an uncontrolled retrospective observational study which describes patient characteristics at presentation, treatment, and response to treatment in 84 adolescents with gender dysphoria, of whom 27 received GnRH analogues. The study used clinical records to show outcomes for up to 13 years (continuing use of GnRH analogues, reasons for stopping GnRH analogues and onward care such as gender-affirming hormone use). The methods are well reported but the results for those taking GnRH analogues are poorly and incompletely reported, particularly for transfemales, and no analysis of data was undertaken. It is difficult to assess the results for stopping GnRH analogues due to incomplete reporting of this outcome.

The sixth study (Klink et al. 2015) was a retrospective longitudinal observational single centre study which assessed bone mineral density in adolescents with gender dysphoria, diagnosed with the DSM-IV-TR criteria. Bone density was assessed when starting GnRH analogues and then when starting gender-affirming hormones. Results are reported for transmales and transfemales separately and no results for the whole cohort are given. Statistical analyses were reported for all outcomes of interest but, because there was no comparator group and participants acted as their own controls, it is not known whether the findings are associated with GnRH analogues or due to changes over time. The authors reported z-scores which allows for comparison with the expected increase in bone density in the general population. However, because no concomitant treatments or comorbidities were

reported it is possible that the findings may not be because of GnRH analogues and there is another way in which the study population differs from the general population.

The seventh study (Schagen et al. 2016) was a prospective observational study of 116 adolescents which provided very low certainty non-comparative evidence on change in serum creatinine between starting GnRH analogues and 1 year, and liver function during treatment. Statistical analyses were reported for changes in serum creatinine but not for liver function. Because there was no comparator group and participants acted as their own controls, it is not known whether the findings are associated with GnRH analogues or due to changes over time, or concomitant treatments.

The eighth study (Staphorsius et al. 2015) was a cross-sectional study of 85 adolescents, 40 with gender dysphoria (of whom 20 were receiving GnRH analogues) and 45 matched controls (not further reported in this evidence review). The study includes 1 outcome of interest for clinical effectiveness (CBCL) and 1 outcome of interest for safety (cognitive development or functioning). The mean (±SD) CBCL, IQ test, reaction time and accuracy scores were given for each group, but the statistical analysis is unclear. It is not reported what analysis was used or which of the groups were compared, therefore it is difficult to interpret the results.

The ninth study (Vlot et al. 2017) was a retrospective observational study which assessed bone mineral apparent density in adolescents with DSM-IV-TR gender dysphoria. Measurements were taken at the start of GnRH analogues and at the start of gender-affirming hormones. Results are reported for young bone age and old bone age in transmales and transfemales separately, and no results for the whole cohort are given. Statistical analyses were reported for all outcomes of interest but, because there was no comparator group and participants acted as their own controls, it is not known whether the findings are associated with GnRH analogues or due to changes over time. The authors reported z-scores which allows for comparison with the expected increase in bone density in the general population. However, because no concomitant treatments or comorbidities were reported it is possible that the findings may not be because of GnRH analogues and there is another way in which the study population differs from the general population.

7. Conclusion

The results of the studies that reported impact on the critical outcomes of gender dysphoria and mental health (depression, anger and anxiety), and the important outcomes of body image and psychosocial impact (global and psychosocial functioning) in children and adolescents with gender dysphoria are of very low certainty using modified GRADE. They suggest little change with GnRH analogues from baseline to follow-up.

Studies that found differences in outcomes could represent changes that are either of questionable clinical value, or the studies themselves are not reliable and changes could be due to confounding, bias or chance. It is plausible, however, that a lack of difference in scores from baseline to follow-up is the effect of GnRH analogues in children and adolescents with gender dysphoria, in whom the development of secondary sexual characteristics might be expected to be associated with an increased impact on gender dysphoria, depression, anxiety, anger and distress over time without treatment. One study reported statistically significant reductions in the Child Behaviour Checklist/Youth Self-Report (CBCL/YSR) scores from

baseline to follow up, and given that the purpose of GnRH analogues is to reduce distress caused by the development of secondary sexual characteristics and the CBCL/YSR in part measures distress, this could be an important finding. However, as the studies all lack reasonable controls not receiving GnRH analogues, the natural history of the outcomes measured in the studies is not known and any positive changes could be a regression to mean.

The results of the studies that reported bone density outcomes suggest that GnRH analogues may reduce the increase in bone density which is expected during puberty. However, as the studies themselves are not reliable, the results could be due to confounding, bias or chance. While controlled trials may not be possible, comparative studies are needed to understand this association and whether the effects of GnRH analogues on bone density are seen after treatment is stopped. All the studies that reported safety outcomes provided very low certainty evidence.

No cost-effectiveness evidence was found to determine whether or not GnRH analogues are cost-effective for children and adolescents with gender dysphoria.

The results of the studies that reported outcomes for subgroups of children and adolescents with gender dysphoria, suggest there may be differences between sex assigned at birth males (transfemales) and sex assigned at birth females (transmales).

Appendix A PICO document

The review questions for this evidence review are:

1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
2. For children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention?
3. For children and adolescents with gender dysphoria, what is the cost-effectiveness of GnRH analogues compared to one or a combination of psychological support, social transitioning to the desired gender or no intervention?
4. From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may derive more (or less) advantage from treatment with GnRH analogues than the wider population of children and adolescents with gender dysphoria?
5. From the evidence selected,
   a) what are the criteria used by the research studies to define gender dysphoria, gender identity disorder and gender incongruence of childhood?
   b) what were the ages at which participants commenced treatment with GnRH analogues?
   c) what was the duration of treatment with GnRH analogues?

PICO table

P – Population and Indication	Children and adolescents aged 18 years or less who have gender dysphoria, gender identity disorder or gender incongruence of childhood as defined by study: The following subgroups of children and adolescents with gender dysphoria, gender identity disorder or gender incongruence of childhood need to be considered: • Sex assigned at birth males. • Sex assigned at birth females. • The duration of gender dysphoria: less than 6 months, 6-24 months, and more than 24 months. • The age of onset of gender dysphoria. • The age at which treatment was initiated. • The age of onset of puberty. • Tanner stage at which treatment was initiated. • Children and adolescents with gender dysphoria who have a pre-existing diagnosis of autistic spectrum disorder. • Children and adolescents with gender dysphoria who had a significant mental health symptom load at diagnosis including anxiety, depression (with or without a history of self-harm and suicidality), suicide attempts, psychosis, personality disorder, Attention Deficit Hyperactivity Disorder and eating disorders.
I – Intervention	Any GnRH analogue including: triptorelin*; buserelin; histrelin; goserelin (Zoladex); leuprorelin/leuprolide (Prostap); nafarelin.

	* Triptorelin (brand names Gonapeptyl and Decapeptyl) are used in Leeds Hospital, England. The search should include brand names as well as generic names.
C – Comparator(s)	One or a combination of: • Psychological support. • Social transitioning to the gender with which the individual identifies. • No intervention.
O – Outcomes	There are no known minimal clinically important differences and there are no preferred timepoints for the outcome measures selected. **All outcomes should be stratified by:** • The age at which treatment with GnRH analogues was initiated. • The length of treatment with GnRH analogues where possible. **A: Clinical Effectiveness** *Critical to decision making* • **Impact on Gender Dysphoria** This outcome is critical because gender dysphoria in adolescents and children is associated with significant distress and problems functioning. Impact on gender dysphoria may be measured by the Utrecht Gender Dysphoria Scale. Other measures as reported in studies may be used as an alternative to the stated measure. • **Impact on mental health** Examples of mental health problems include self-harm, thoughts of suicide, suicide attempts, eating disorders, depression/low mood and anxiety. These outcomes are critical because self-harm and thoughts of suicide have the potential to result in significant physical harm and for completed suicides the death of the young person. Disordered eating habits may cause significant morbidity in young people. Depression and anxiety are also critical outcomes because they may impact on social, occupational, or other areas of functioning of children and adolescents. The Child and Adolescent Psychiatric Assessment (CAPA) may be used to measure depression and anxiety. The impact on self-harm and suicidality (ideation and behaviour) may be measured using the Suicide Ideation Questionnaire Junior. Other measures may be used as an alternative to the stated measures. • **Impact on Quality of Life** This outcome is critical because gender dysphoria in children and adolescents may be associated with a significant reduction in health-related quality of life. Quality of Life may be measured by the KINDL questionnaire, Kidscreen 52. Other measures as reported in studies may be used as an alternative to the stated measure. *Important to decision making* • **Impact on body Image** This outcome is important because some transgender young people may desire to take steps to suppress features of their physical appearance associated with their sex assigned at birth or accentuate physical features of their desired gender. The Body Image Scale could be used as a measure. Other measures

	as reported in studies may also be used as an alternative to the stated measure. • **Psychosocial Impact** Examples of psychosocial impact are: coping mechanisms which may impact on substance misuse; family relationships; peer relationships. This outcome is important because gender dysphoria in adolescents and children is associated with internalising and externalising behaviours and emotional and behavioural problems which may impact on social and occupational functioning. The child behavioural check list (CBCL) may be used to measure the impact on psychosocial functioning. Other measures as reported in studies may be used as an alternative to the stated measure. • **Engagement with health care services** This outcome is important because patient engagement with healthcare services will impact on their clinical outcomes. Engagement with health care services may be measured using the Youth Health Care measure-satisfaction, utilization, and needs (YHC-SUN) questionnaire. Loss to follow up should also be ascertained as part of this outcome. Alternative measures to the YHC-SUN questionnaire may be used as reported in studies. • **Transitioning surgery – Impact on extent of and satisfaction with surgery** This outcome is important because some children and adolescents with gender dysphoria may proceed to transitioning surgery. Stated measures of the extent of transitioning surgery and satisfaction with surgery in studies may be reported. • **Stopping treatment** The proportion of patients who stop treatment with GnRH analogues and the reasons why. This outcome is important to patients because there is uncertainty about the short- and long-term safety and adverse effects of GnRH analogues in children and adolescents being treated for gender dysphoria. **B: Safety** • Short and long-term safety and adverse effects of taking GnRH analogues are important because GnRH analogues are not licensed for the treatment of adolescents and children with gender dysphoria. Aspects to be reported on should include: o Impact of the drug use such as its impact on bone density, arterial hypertension, cognitive development/functioning o Impact of withdrawing the drug such as, slipped upper femoral epiphysis, reversibility on the reproductive system, and any others as reported. **C: Cost effectiveness** Cost effectiveness studies should be reported.
Inclusion criteria	
Study design	Systematic reviews, randomised controlled trials, controlled clinical trials, cohort studies. If no higher level quality evidence is found, case series can be considered.

Language	English only
Patients	Human studies only
Age	18 years or less
Date limits	2000-2020
Exclusion criteria	
Publication type	Conference abstracts, non-systematic reviews, narrative reviews, commentaries, letters, editorials, guidelines and pre-publication prints
Study design	Case reports, resource utilisation studies

Appendix B Search strategy

Medline, Embase, the Cochrane Library, HTA and APA PsycInfo were searched on 23 July 2020, limiting the search to papers published in English language in the last 20 years. Conference abstracts and letters were excluded.

Database: Medline
Platform: Ovid
Version: Ovid MEDLINE(R) <1946 to July 21, 2020>
Search date: 23/7/2020
Number of results retrieved: 144
Search strategy:

1 Gender Dysphoria/ (485)
2 Gender Identity/ (18452)
3 "Sexual and Gender Disorders"/ (75)
4 Transsexualism/ (3758)
5 Transgender Persons/ (3143)
6 Health Services for Transgender Persons/ (136)
7 exp Sex Reassignment Procedures/ (836)
8 (gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (7435)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (12678)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (102343)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (6974)
12 (male-to-female or m2f or female-to-male or f2m).tw. (114841)
13 or/1-12 (252702)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (1137479)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (852400)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (1913257)

17 Minors/ (2574)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (2361686)
19 exp pediatrics/ (58118)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (836269)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (2024207)
22 Puberty/ (13278)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (424246)
24 Schools/ (38104)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (7199)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (468992)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (89353)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (887838)
29 or/14-28 (5534171)
30 13 and 29 (79263)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (7)
32 30 or 31 (79263)
33 Gonadotropin-Releasing Hormone/ (27588)
34 (pubert* adj3 block*).ti,ab. (78)
35 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (17299)
36 (GnRH adj2 analog*).ti,ab. (2541)
37 GnRH*.ti,ab. (20991)
38 "GnRH agonist*".ti,ab. (4040)
39 Triptorelin Pamoate/ (1906)
40 triptorelin.ti,ab. (677)
41 arvekap.ti,ab. (1)
42 ("AY 25650" or AY25650).ti,ab. (1)
43 ("BIM 21003" or BIM21003).ti,ab. (0)
44 ("BN 52014" or BN52014).ti,ab. (0)
45 ("CL 118532" or CL118532).ti,ab. (0)
46 Debio.ti,ab. (83)
47 diphereline.ti,ab. (17)
48 moapar.ti,ab. (0)
49 pamorelin.ti,ab. (0)
50 trelstar.ti,ab. (3)
51 triptodur.ti,ab. (1)
52 ("WY 42422" or WY42422).ti,ab. (0)
53 ("WY 42462" or WY42462).ti,ab. (0)
54 gonapeptyl.ti,ab. (0)
55 decapeptyl.ti,ab. (210)
56 salvacyl.ti,ab. (0)
57 Buserelin/ (2119)
58 buserelin.ti,ab. (1304)

59 bigonist.ti,ab. (0)
60 ("hoe 766" or hoe-766 or hoe766).ti,ab. (69)
61 profact.ti,ab. (2)
62 receptal.ti,ab. (30)
63 suprecur.ti,ab. (4)
64 suprefact.ti,ab. (22)
65 tiloryth.ti,ab. (0)
66 histrelin.ti,ab. (55)
67 "LHRH-hydrogel implant".ti,ab. (1)
68 ("RL 0903" or RL0903).ti,ab. (1)
69 ("SPD 424" or SPD424).ti,ab. (1)
70 goserelin.ti,ab. (875)
71 Goserelin/ (1612)
72 ("ici 118630" or ici118630).ti,ab. (51)
73 ("ZD-9393" or ZD9393).ti,ab. (0)
74 zoladex.ti,ab. (379)
75 leuprorelin.ti,ab. (413)
76 carcinil.ti,ab. (0)
77 enanton*.ti,ab. (23)
78 ginecrin.ti,ab. (0)
79 leuplin.ti,ab. (13)
80 Leuprolide/ (2900)
81 leuprolide.ti,ab. (1743)
82 lucrin.ti,ab. (11)
83 lupron.ti,ab. (162)
84 provren.ti,ab. (0)
85 procrin.ti,ab. (3)
86 ("tap 144" or tap144).ti,ab. (40)
87 (a-43818 or a43818).ti,ab. (3)
88 Trenantone.ti,ab. (1)
89 staladex.ti,ab. (0)
90 prostap.ti,ab. (6)
91 Nafarelin/ (327)
92 nafarelin.ti,ab. (251)
93 ("76932-56-4" or "76932564").ti,ab. (0)
94 ("76932-60-0" or "76932600").ti,ab. (0)
95 ("86220-42-0" or "86220420").ti,ab. (0)
96 ("rs 94991 298" or rs94991298).ti,ab. (0)
97 synarel.ti,ab. (12)
98 deslorelin.ti,ab. (263)
99 gonadorelin.ti,ab. (201)
100 ("33515-09-2" or "33515092").ti,ab. (0)
101 ("51952-41-1" or "51952411").ti,ab. (0)
102 ("52699-48-6" or "52699486").ti,ab. (0)
103 cetrorelix.ti,ab. (463)
104 cetrotide.ti,ab. (41)
105 ("NS 75A" or NS75A).ti,ab. (0)
106 ("NS 75B" or NS75B).ti,ab. (0)

107 ("SB 075" or SB075).ti,ab. (0)
108 ("SB 75" or SB75).ti,ab. (63)
109 gonadoliberin.ti,ab. (143)
110 kryptocur.ti,ab. (6)
111 cetrorelix.ti,ab. (463)
112 cetrotide.ti,ab. (41)
113 antagon.ti,ab. (17)
114 ganirelix.ti,ab. (138)
115 ("ORG 37462" or ORG37462).ti,ab. (3)
116 orgalutran.ti,ab. (20)
117 ("RS 26306" or RS26306).ti,ab. (5)
118 ("AY 24031" or AY24031).ti,ab. (0)
119 factrel.ti,ab. (11)
120 fertagyl.ti,ab. (11)
121 lutrelef.ti,ab. (5)
122 lutrepulse.ti,ab. (3)
123 relefact.ti,ab. (10)
124 fertiral.ti,ab. (0)
125 (hoe471 or "hoe 471").ti,ab. (6)
126 relisorm.ti,ab. (4)
127 cystorelin.ti,ab. (18)
128 dirigestran.ti,ab. (5)
129 or/33-128 (42216)
130 32 and 129 (416)
131 limit 130 to english language (393)
132 limit 131 to (letter or historical article or comment or editorial or news or case reports) (36)
133 131 not 132 (357)
134 animals/ not humans/ (4686361)
135 133 not 134 (181)
136 limit 135 to yr="2000 -Current" (144)

Database: Medline in-process
Platform: Ovid
Version: Ovid MEDLINE(R) In-Process & Other Non-Indexed Citations <1946 to July 21, 2020>
Search date: 23/7/2020
Number of results retrieved:
Search strategy: 42

1 Gender Dysphoria/ (0)
2 Gender Identity/ (0)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (0)
5 Transgender Persons/ (0)
6 Health Services for Transgender Persons/ (0)
7 exp Sex Reassignment Procedures/ (0)

8 (gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (1645)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (2333)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (20884)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (968)
12 (male-to-female or m2f or female-to-male or f2m).tw. (15513)
13 or/1-12 (39905)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (0)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (80723)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (0)
17 Minors/ (0)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (321871)
19 exp pediatrics/ (0)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (119783)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (0)
22 Puberty/ (0)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (60264)
24 Schools/ (0)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (0)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (69233)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (10319)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (112800)
29 or/14-28 (525529)
30 13 and 29 (9196)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (3)
32 30 or 31 (9197)
33 Gonadotropin-Releasing Hormone/ (0)
34 (pubert* adj3 block*).ti,ab. (19)
35 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (1425)
36 (GnRH adj2 analog*).ti,ab. (183)
37 GnRH*.ti,ab. (1695)
38 "GnRH agonist*".ti,ab. (379)
39 Triptorelin Pamoate/ (0)
40 triptorelin.ti,ab. (72)
41 arvekap.ti,ab. (0)
42 ("AY 25650" or AY25650).ti,ab. (0)
43 ("BIM 21003" or BIM21003).ti,ab. (0)
44 ("BN 52014" or BN52014).ti,ab. (0)
45 ("CL 118532" or CL118532).ti,ab. (0)

46 Debio.ti,ab. (11)
47 diphereline.ti,ab. (6)
48 moapar.ti,ab. (0)
49 pamorelin.ti,ab. (0)
50 trelstar.ti,ab. (0)
51 triptodur.ti,ab. (0)
52 ("WY 42422" or WY42422).ti,ab. (0)
53 ("WY 42462" or WY42462).ti,ab. (0)
54 gonapeptyl.ti,ab. (0)
55 decapeptyl.ti,ab. (8)
56 salvacyl.ti,ab. (0)
57 Buserelin/ (0)
58 buserelin.ti,ab. (59)
59 bigonist.ti,ab. (0)
60 ("hoe 766" or hoe-766 or hoe766).ti,ab. (3)
61 profact.ti,ab. (0)
62 receptal.ti,ab. (0)
63 suprecur.ti,ab. (1)
64 suprefact.ti,ab. (2)
65 tiloryth.ti,ab. (0)
66 histrelin.ti,ab. (9)
67 "LHRH-hydrogel implant".ti,ab. (0)
68 ("RL 0903" or RL0903).ti,ab. (0)
69 ("SPD 424" or SPD424).ti,ab. (0)
70 goserelin.ti,ab. (68)
71 Goserelin/ (0)
72 ("ici 118630" or ici118630).ti,ab. (0)
73 ("ZD-9393" or ZD9393).ti,ab. (0)
74 zoladex.ti,ab. (6)
75 leuprorelin.ti,ab. (47)
76 carcinil.ti,ab. (0)
77 enanton*.ti,ab. (1)
78 ginecrin.ti,ab. (0)
79 leuplin.ti,ab. (1)
80 Leuprolide/ (0)
81 leuprolide.ti,ab. (121)
82 lucrin.ti,ab. (4)
83 lupron.ti,ab. (10)
84 provren.ti,ab. (0)
85 procrin.ti,ab. (0)
86 ("tap 144" or tap144).ti,ab. (0)
87 (a-43818 or a43818).ti,ab. (0)
88 Trenantone.ti,ab. (1)
89 staladex.ti,ab. (0)
90 prostap.ti,ab. (0)
91 Nafarelin/ (0)
92 nafarelin.ti,ab. (5)
93 ("76932-56-4" or "76932564").ti,ab. (0)

94 ("76932-60-0" or "76932600").ti,ab. (0)
95 ("86220-42-0" or "86220420").ti,ab. (0)
96 ("rs 94991 298" or rs94991298).ti,ab. (0)
97 synarel.ti,ab. (0)
98 deslorelin.ti,ab. (14)
99 gonadorelin.ti,ab. (13)
100 ("33515-09-2" or "33515092").ti,ab. (0)
101 ("51952-41-1" or "51952411").ti,ab. (0)
102 ("52699-48-6" or "52699486").ti,ab. (0)
103 cetrorelix.ti,ab. (31)
104 cetrotide.ti,ab. (5)
105 ("NS 75A" or NS75A).ti,ab. (0)
106 ("NS 75B" or NS75B).ti,ab. (0)
107 ("SB 075" or SB075).ti,ab. (0)
108 ("SB 75" or SB75).ti,ab. (2)
109 gonadoliberin.ti,ab. (4)
110 kryptocur.ti,ab. (1)
111 cetrorelix.ti,ab. (31)
112 cetrotide.ti,ab. (5)
113 antagon.ti,ab. (0)
114 ganirelix.ti,ab. (8)
115 ("ORG 37462" or ORG37462).ti,ab. (0)
116 orgalutran.ti,ab. (3)
117 ("RS 26306" or RS26306).ti,ab. (0)
118 ("AY 24031" or AY24031).ti,ab. (0)
119 factrel.ti,ab. (2)
120 fertagyl.ti,ab. (1)
121 lutrelef.ti,ab. (0)
122 lutrepulse.ti,ab. (0)
123 relefact.ti,ab. (0)
124 fertiral.ti,ab. (0)
125 (hoe471 or "hoe 471").ti,ab. (0)
126 relisorm.ti,ab. (0)
127 cystorelin.ti,ab. (1)
128 dirigestran.ti,ab. (0)
129 or/33-128 (2332)
130 32 and 129 (45)
131 limit 130 to english language (45)
132 limit 131 to yr="2000 -Current" (42)

Database: Medline epubs ahead of print
Platform: Ovid
Version: Ovid MEDLINE(R) Epub Ahead of Print <July 21, 2020>
Search date: 23/7/2020
Number of results retrieved: 8
Search strategy:

1 Gender Dysphoria/ (0)

2 Gender Identity/ (0)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (0)
5 Transgender Persons/ (0)
6 Health Services for Transgender Persons/ (0)
7 exp Sex Reassignment Procedures/ (0)
8 (gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (486)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (640)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (1505)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (178)
12 (male-to-female or m2f or female-to-male or f2m).tw. (2480)
13 or/1-12 (4929)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (0)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (15496)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (0)
17 Minors/ (0)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (53563)
19 exp pediatrics/ (0)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (22796)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (0)
22 Puberty/ (0)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (13087)
24 Schools/ (0)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (0)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (12443)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (1416)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (20166)
29 or/14-28 (88366)
30 13 and 29 (1638)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (1)
32 30 or 31 (1638)
33 Gonadotropin-Releasing Hormone/ (0)
34 (pubert* adj3 block*).ti,ab. (2)
35 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (176)
36 (GnRH adj2 analog*).ti,ab. (30)
37 GnRH*.ti,ab. (223)
38 "GnRH agonist*".ti,ab. (49)
39 Triptorelin Pamoate/ (0)

40 triptorelin.ti,ab. (12)
41 arvekap.ti,ab. (0)
42 ("AY 25650" or AY25650).ti,ab. (0)
43 ("BIM 21003" or BIM21003).ti,ab. (0)
44 ("BN 52014" or BN52014).ti,ab. (0)
45 ("CL 118532" or CL118532).ti,ab. (0)
46 Debio.ti,ab. (2)
47 diphereline.ti,ab. (1)
48 moapar.ti,ab. (0)
49 pamorelin.ti,ab. (0)
50 trelstar.ti,ab. (0)
51 triptodur.ti,ab. (0)
52 ("WY 42422" or WY42422).ti,ab. (0)
53 ("WY 42462" or WY42462).ti,ab. (0)
54 gonapeptyl.ti,ab. (0)
55 decapeptyl.ti,ab. (0)
56 salvacyl.ti,ab. (0)
57 Buserelin/ (0)
58 buserelin.ti,ab. (7)
59 bigonist.ti,ab. (0)
60 ("hoe 766" or hoe-766 or hoe766).ti,ab. (0)
61 profact.ti,ab. (0)
62 receptal.ti,ab. (0)
63 suprecur.ti,ab. (0)
64 suprefact.ti,ab. (1)
65 tiloryth.ti,ab. (0)
66 histrelin.ti,ab. (2)
67 "LHRH-hydrogel implant".ti,ab. (0)
68 ("RL 0903" or RL0903).ti,ab. (0)
69 ("SPD 424" or SPD424).ti,ab. (0)
70 goserelin.ti,ab. (11)
71 Goserelin/ (0)
72 ("ici 118630" or ici118630).ti,ab. (0)
73 ("ZD-9393" or ZD9393).ti,ab. (0)
74 zoladex.ti,ab. (1)
75 leuprorelin.ti,ab. (13)
76 carcinil.ti,ab. (0)
77 enanton*.ti,ab. (1)
78 ginecrin.ti,ab. (0)
79 leuplin.ti,ab. (0)
80 Leuprolide/ (0)
81 leuprolide.ti,ab. (22)
82 lucrin.ti,ab. (0)
83 lupron.ti,ab. (2)
84 provren.ti,ab. (0)
85 procrin.ti,ab. (0)
86 ("tap 144" or tap144).ti,ab. (1)
87 (a-43818 or a43818).ti,ab. (0)

88 Trenantone.ti,ab. (0)
89 staladex.ti,ab. (0)
90 prostap.ti,ab. (0)
91 Nafarelin/ (0)
92 nafarelin.ti,ab. (4)
93 ("76932-56-4" or "76932564").ti,ab. (0)
94 ("76932-60-0" or "76932600").ti,ab. (0)
95 ("86220-42-0" or "86220420").ti,ab. (0)
96 ("rs 94991 298" or rs94991298).ti,ab. (0)
97 synarel.ti,ab. (0)
98 deslorelin.ti,ab. (3)
99 gonadorelin.ti,ab. (3)
100 ("33515-09-2" or "33515092").ti,ab. (0)
101 ("51952-41-1" or "51952411").ti,ab. (0)
102 ("52699-48-6" or "52699486").ti,ab. (0)
103 cetrorelix.ti,ab. (6)
104 cetrotide.ti,ab. (2)
105 ("NS 75A" or NS75A).ti,ab. (0)
106 ("NS 75B" or NS75B).ti,ab. (0)
107 ("SB 075" or SB075).ti,ab. (0)
108 ("SB 75" or SB75).ti,ab. (0)
109 gonadoliberin.ti,ab. (0)
110 kryptocur.ti,ab. (0)
111 cetrorelix.ti,ab. (6)
112 cetrotide.ti,ab. (2)
113 antagon.ti,ab. (1)
114 ganirelix.ti,ab. (1)
115 ("ORG 37462" or ORG37462).ti,ab. (0)
116 orgalutran.ti,ab. (0)
117 ("RS 26306" or RS26306).ti,ab. (0)
118 ("AY 24031" or AY24031).ti,ab. (0)
119 factrel.ti,ab. (0)
120 fertagyl.ti,ab. (0)
121 lutrelef.ti,ab. (0)
122 lutrepulse.ti,ab. (0)
123 relefact.ti,ab. (0)
124 fertiral.ti,ab. (0)
125 (hoe471 or "hoe 471").ti,ab. (0)
126 relisorm.ti,ab. (0)
127 cystorelin.ti,ab. (0)
128 dirigestran.ti,ab. (0)
129 or/33-128 (310)
130 32 and 129 (8)
131 limit 130 to english language (8)
132 limit 131 to yr="2000 -Current" (8)

Database: Medline daily update
Platform: Ovid

Version: Ovid MEDLINE(R) Daily Update <July 21, 2020>
Search date: 23/7/2020
Number of results retrieved: 1
Search strategy

1 Gender Dysphoria/ (4)
2 Gender Identity/ (38)
3 "Sexual and Gender Disorders"/ (0)
4 Transsexualism/ (2)
5 Transgender Persons/ (26)
6 Health Services for Transgender Persons/ (1)
7 exp Sex Reassignment Procedures/ (3)
8 (gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (24)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (39)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (87)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (15)
12 (male-to-female or m2f or female-to-male or f2m).tw. (181)
13 or/1-12 (358)
14 exp Infant/ or Infant Health/ or Infant Welfare/ (932)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (981)
16 exp Child/ or exp Child Behavior/ or Child Health/ or Child Welfare/ (1756)
17 Minors/ (3)
18 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (3672)
19 exp pediatrics/ (75)
20 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (1658)
21 Adolescent/ or Adolescent Behavior/ or Adolescent Health/ (2006)
22 Puberty/ (8)
23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (732)
24 Schools/ (56)
25 Child Day Care Centers/ or exp Nurseries/ or Schools, Nursery/ (5)
26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (622)
27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (98)
28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (1301)
29 or/14-28 (6705)
30 13 and 29 (130)
31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (0)
32 30 or 31 (130)
33 Gonadotropin-Releasing Hormone/ (11)

34 (pubert* adj3 block*).ti,ab. (0)
35 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (10)
36 (GnRH adj2 analog*).ti,ab. (2)
37 GnRH*.ti,ab. (14)
38 "GnRH agonist*".ti,ab. (4)
39 Triptorelin Pamoate/ (1)
40 triptorelin.ti,ab. (1)
41 arvekap.ti,ab. (0)
42 ("AY 25650" or AY25650).ti,ab. (0)
43 ("BIM 21003" or BIM21003).ti,ab. (0)
44 ("BN 52014" or BN52014).ti,ab. (0)
45 ("CL 118532" or CL118532).ti,ab. (0)
46 Debio.ti,ab. (1)
47 diphereline.ti,ab. (0)
48 moapar.ti,ab. (0)
49 pamorelin.ti,ab. (0)
50 trelstar.ti,ab. (0)
51 triptodur.ti,ab. (0)
52 ("WY 42422" or WY42422).ti,ab. (0)
53 ("WY 42462" or WY42462).ti,ab. (0)
54 gonapeptyl.ti,ab. (0)
55 decapeptyl.ti,ab. (0)
56 salvacyl.ti,ab. (0)
57 Buserelin/ (0)
58 buserelin.ti,ab. (0)
59 bigonist.ti,ab. (0)
60 ("hoe 766" or hoe-766 or hoe766).ti,ab. (0)
61 profact.ti,ab. (0)
62 receptal.ti,ab. (0)
63 suprecur.ti,ab. (0)
64 suprefact.ti,ab. (0)
65 tiloryth.ti,ab. (0)
66 histrelin.ti,ab. (0)
67 "LHRH-hydrogel implant".ti,ab. (0)
68 ("RL 0903" or RL0903).ti,ab. (0)
69 ("SPD 424" or SPD424).ti,ab. (0)
70 goserelin.ti,ab. (1)
71 Goserelin/ (2)
72 ("ici 118630" or ici118630).ti,ab. (0)
73 ("ZD-9393" or ZD9393).ti,ab. (0)
74 zoladex.ti,ab. (0)
75 leuprorelin.ti,ab. (0)
76 carcinil.ti,ab. (0)
77 enanton*.ti,ab. (0)
78 ginecrin.ti,ab. (0)
79 leuplin.ti,ab. (0)
80 Leuprolide/ (0)
81 leuprolide.ti,ab. (0)

82 lucrin.ti,ab. (0)
83 lupron.ti,ab. (0)
84 provren.ti,ab. (0)
85 procrin.ti,ab. (0)
86 ("tap 144" or tap144).ti,ab. (0)
87 (a-43818 or a43818).ti,ab. (0)
88 Trenantone.ti,ab. (0)
89 staladex.ti,ab. (0)
90 prostap.ti,ab. (0)
91 Nafarelin/ (0)
92 nafarelin.ti,ab. (0)
93 ("76932-56-4" or "76932564").ti,ab. (0)
94 ("76932-60-0" or "76932600").ti,ab. (0)
95 ("86220-42-0" or "86220420").ti,ab. (0)
96 ("rs 94991 298" or rs94991298).ti,ab. (0)
97 synarel.ti,ab. (0)
98 deslorelin.ti,ab. (0)
99 gonadorelin.ti,ab. (0)
100 ("33515-09-2" or "33515092").ti,ab. (0)
101 ("51952-41-1" or "51952411").ti,ab. (0)
102 ("52699-48-6" or "52699486").ti,ab. (0)
103 cetrorelix.ti,ab. (0)
104 cetrotide.ti,ab. (0)
105 ("NS 75A" or NS75A).ti,ab. (0)
106 ("NS 75B" or NS75B).ti,ab. (0)
107 ("SB 075" or SB075).ti,ab. (0)
108 ("SB 75" or SB75).ti,ab. (0)
109 gonadoliberin.ti,ab. (0)
110 kryptocur.ti,ab. (0)
111 cetrorelix.ti,ab. (0)
112 cetrotide.ti,ab. (0)
113 antagon.ti,ab. (0)
114 ganirelix.ti,ab. (0)
115 ("ORG 37462" or ORG37462).ti,ab. (0)
116 orgalutran.ti,ab. (0)
117 ("RS 26306" or RS26306).ti,ab. (0)
118 ("AY 24031" or AY24031).ti,ab. (0)
119 factrel.ti,ab. (0)
120 fertagyl.ti,ab. (0)
121 lutrelef.ti,ab. (0)
122 lutrepulse.ti,ab. (0)
123 relefact.ti,ab. (0)
124 fertiral.ti,ab. (0)
125 (hoe471 or "hoe 471").ti,ab. (0)
126 relisorm.ti,ab. (0)
127 cystorelin.ti,ab. (0)
128 dirigestran.ti,ab. (0)
129 or/33-128 (23)

130 32 and 129 (1)
131 limit 130 to english language (1)
132 limit 131 to yr="2000 -Current" (1)

Database: Embase
Platform: Ovid
Version: Embase <1974 to 2020 July 22>
Search date: 23/7/2020
Number of results retrieved: 367
Search strategy:

1 exp Gender Dysphoria/ (5399)
2 Gender Identity/ (16820)
3 "Sexual and Gender Disorders"/ (24689)
4 Transsexualism/ (3869)
5 exp Transgender/ (6597)
6 Health Services for Transgender Persons/ (158848)
7 exp Sex Reassignment Procedures/ or sex transformation/ (3058)
8 (gender* adj3 (dysphori* or affirm* or incongru* or identi* or disorder* or confus* or minorit* or queer*)).tw. (13005)
9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (22509)
10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (154446)
11 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (10327)
12 (male-to-female or m2f or female-to-male or f2m).tw. (200166)
13 or/1-12 (582812)
14 exp juvenile/ or Child Behavior/ or Child Welfare/ or Child Health/ or infant welfare/ or "minor (person)"/ or elementary student/ (3437324)
15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (1186161)
16 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (3586795)
17 exp pediatrics/ (106214)
18 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (1491597)
19 exp adolescence/ or exp adolescent behavior/ or adolescent health/ or high school student/ or middle school student/ (105108)
20 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (641660)
21 school/ or high school/ or kindergarten/ or middle school/ or primary school/ or nursery school/ or day care/ (103791)
22 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (687437)
23 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (138908)
24 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (1562903)

25 or/14-24 (7130881)
26 13 and 25 (182161)
27 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (17)
28 26 or 27 (182161)
29 gonadorelin/ (37580)
30 (pubert* adj3 block*).ti,ab. (142)
31 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (21450)
32 (GnRH adj2 analog*).ti,ab. (4013)
33 GnRH*.ti,ab. (29862)
34 "GnRH agonist*".ti,ab. (6719)
35 exp gonadorelin agonist/ or gonadorelin derivative/ or gonadorelin acetate/ (23304)
36 Triptorelin/ (5427)
37 triptorelin.ti,ab. (1182)
38 arvekap.ti,ab. (3)
39 ("AY 25650" or AY25650).ti,ab. (1)
40 ("BIM 21003" or BIM21003).ti,ab. (0)
41 ("BN 52014" or BN52014).ti,ab. (0)
42 ("CL 118532" or CL118532).ti,ab. (0)
43 Debio.ti,ab. (185)
44 diphereline.ti,ab. (51)
45 moapar.ti,ab. (0)
46 pamorelin.ti,ab. (0)
47 trelstar.ti,ab. (5)
48 triptodur.ti,ab. (1)
49 ("WY 42422" or WY42422).ti,ab. (0)
50 ("WY 42462" or WY42462).ti,ab. (0)
51 gonapeptyl.ti,ab. (10)
52 decapeptyl.ti,ab. (307)
53 salvacyl.ti,ab. (1)
54 buserelin acetate/ or buserelin/ (5164)
55 buserelin.ti,ab. (1604)
56 bigonist.ti,ab. (1)
57 ("hoe 766" or hoe-766 or hoe766).ti,ab. (89)
58 profact.ti,ab. (4)
59 receptal.ti,ab. (37)
60 suprecur.ti,ab. (8)
61 suprefact.ti,ab. (30)
62 tiloryth.ti,ab. (0)
63 histrelin/ (446)
64 histrelin.ti,ab. (107)
65 "LHRH-hydrogel implant".ti,ab. (1)
66 ("RL 0903" or RL0903).ti,ab. (1)
67 ("SPD 424" or SPD424).ti,ab. (1)
68 goserelin.ti,ab. (1487)
69 Goserelin/ (7128)
70 ("ici 118630" or ici118630).ti,ab. (49)
71 ("ZD-9393" or ZD9393).ti,ab. (0)

72 zoladex.ti,ab. (501)
73 leuprorelin/ (11312)
74 leuprorelin.ti,ab. (727)
75 carcinil.ti,ab. (0)
76 enanton*.ti,ab. (38)
77 ginecrin.ti,ab. (1)
78 leuplin.ti,ab. (26)
79 leuprolide.ti,ab. (2788)
80 lucrin.ti,ab. (47)
81 lupron.ti,ab. (361)
82 provren.ti,ab. (0)
83 procrin.ti,ab. (11)
84 ("tap 144" or tap144).ti,ab. (63)
85 (a-43818 or a43818).ti,ab. (3)
86 Trenantone.ti,ab. (7)
87 staladex.ti,ab. (0)
88 prostap.ti,ab. (11)
89 nafarelin acetate/ or nafarelin/ (1441)
90 nafarelin.ti,ab. (324)
91 ("76932-56-4" or "76932564").ti,ab. (0)
92 ("76932-60-0" or "76932600").ti,ab. (0)
93 ("86220-42-0" or "86220420").ti,ab. (0)
94 ("rs 94991 298" or rs94991298).ti,ab. (0)
95 synarel.ti,ab. (28)
96 deslorelin/ (452)
97 deslorelin.ti,ab. (324)
98 gonadorelin.ti,ab. (338)
99 ("33515-09-2" or "33515092").ti,ab. (0)
100 ("51952-41-1" or "51952411").ti,ab. (0)
101 ("52699-48-6" or "52699486").ti,ab. (0)
102 cetrorelix/ (2278)
103 cetrorelix.ti,ab. (717)
104 cetrotide.ti,ab. (113)
105 ("NS 75A" or NS75A).ti,ab. (0)
106 ("NS 75B" or NS75B).ti,ab. (0)
107 ("SB 075" or SB075).ti,ab. (1)
108 ("SB 75" or SB75).ti,ab. (76)
109 gonadoliberin.ti,ab. (152)
110 kryptocur.ti,ab. (6)
111 cetrorelix.ti,ab. (717)
112 cetrotide.ti,ab. (113)
113 antagon.ti,ab. (32)
114 ganirelix/ (1284)
115 ganirelix.ti,ab. (293)
116 ("ORG 37462" or ORG37462).ti,ab. (4)
117 orgalutran/ (1284)
118 orgalutran.ti,ab. (68)
119 ("RS 26306" or RS26306).ti,ab. (6)

120 ("AY 24031" or AY24031).ti,ab. (0)
121 factrel.ti,ab. (14)
122 fertagyl.ti,ab. (20)
123 lutrelef.ti,ab. (7)
124 lutrepulse.ti,ab. (6)
125 relefact.ti,ab. (10)
126 fertiral.ti,ab. (0)
127 (hoe471 or "hoe 471").ti,ab. (4)
128 relisorm.ti,ab. (6)
129 cystorelin.ti,ab. (26)
130 dirigestran.ti,ab. (5)
131 or/29-130 (80790)
132 28 and 131 (988)
133 limit 132 to english language (940)
134 133 not (letter or editorial).pt. (924)
135 134 not (conference abstract or conference paper or conference proceeding or "conference review").pt. (683)
136 nonhuman/ not (human/ and nonhuman/) (4649157)
137 135 not 136 (506)
138 limit 137 to yr="2000 -Current" (420)
139 elsevier.cr. (25912990)
140 138 and 139 (372)
141 remove duplicates from 140 (367)

Database: Cochrane Library – incorporating Cochrane Database of Systematic Reviews (CDSR); CENTRAL
Platform: Wiley
Version:
CDSR – Issue 7 of 12, July 2020
CENTRAL – Issue 7 of 12, July 2020
Search date: 23/7/2020
Number of results retrieved: CDSR – 1; CENTRAL - 8.

#1 [mh ^"Gender Dysphoria"] 3
#2 [mh ^"gender identity"] 227
#3 [mh ^"sexual and gender disorders"] 2
#4 [mh ^transsexualism] 27
#5 [mh ^"transgender persons"] 36
#6 [mh ^"health services for transgender persons"] 0
#7 [mh "sex reassignment procedures"] 4
#8 (gender* NEAR/3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)):ti,ab 308
#9 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*):ti,ab 929
#10 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*):ti,ab 3915
#11 ((sex or gender*) NEAR/3 (reassign* or chang* or transform* or transition*)):ti,ab 493
#12 (male-to-female or m2f or female-to-male or f2m):ti,ab 489

#13 {or #1-#12} 6142
#14 [mh infant] or [mh ^"infant health"] or [mh ^"infant welfare"] 27769
#15 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*):ti,ab 69476
#16 [mh child] or [mh "child behavior"] or [mh ^"child health"] or [mh ^"child welfare"] 42703
#17 [mh ^minors] 8
#18 (child* or minor or minors or boy* or girl* or kid or kids or young*):ti,ab 175826
#19 [mh pediatrics]661
#20 (pediatric* or paediatric* or peadiatric*):ti,ab 30663
#21 [mh ^adolescent] or [mh ^"adolescent behavior"] or [mh ^"adolescent health"] 102154
#22 [mh ^puberty] 295
#23 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*):ti,ab 34139
#24 [mh ^schools] 1914
#25 [mh ^"Child Day Care Centers"] or [mh nurseries] or [mh ^"schools, nursery"] 277
#26 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*):ti,ab 54723
#27 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") NEAR/2 (year or years or age or ages or aged)):ti,ab 6710
#28 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") NEAR/2 (year or years or age or ages or aged)):ti,ab 196881
#29 {or #14-#28} 469351
#30 #13 and #29 2146
#31 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*):ti,ab 0
#32 #30 or #31 2146
#33 [mh ^"Gonadotropin-Releasing Hormone"] 1311
#34 (pubert* NEAR/3 block*):ti,ab 1
#35 ((gonadotrophin or gonadotropin) and releasing):ti,ab 2095
#36 (GnRH NEAR/2 analog*):ti,ab 493
#37 GnRH*:ti,ab 3764
#38 "GnRH agonist*":ti,ab 1399
#39 [mh ^"Triptorelin Pamoate"] 451
#40 triptorelin:ti,ab 451
#41 arvekap:ti,ab 4
#42 ("AY 25650" or AY25650):ti,ab 0
#43 ("BIM 21003" or BIM21003):ti,ab 0
#44 ("BN 52014" or BN52014):ti,ab 0
#45 ("CL 118532" or CL118532):ti,ab 0
#46 Debio:ti,ab 301
#47 diphereline:ti,ab 25
#48 moapar:ti,ab 0
#49 pamorelin:ti,ab 5
#50 trelstar:ti,ab 3

#51 triptodur:ti,ab 0
#52 ("WY 42422" or WY42422):ti,ab 0
#53 ("WY 42462" or WY42462):ti,ab 0
#54 gonapeptyl:ti,ab 11
#55 decapeptyl:ti,ab 135
#56 salvacyl:ti,ab 0
#57 [mh ^Buserelin] 290
#58 Buserelin:ti,ab 339
#59 bigonist:ti,ab 0
#60 ("hoe 766" or hoe-766 or hoe766):ti,ab 11
#61 profact:ti,ab 1
#62 receptal:ti,ab 4
#63 suprecur:ti,ab 0
#64 suprefact:ti,ab 28
#65 tiloryth:ti,ab 0
#66 histrelin:ti,ab 5
#67 "LHRH-hydrogel implant":ti,ab 0
#68 ("RL 0903" or RL0903):ti,ab 0
#69 ("SPD 424" or SPD424):ti,ab 0
#70 goserelin:ti,ab 761
#71 [mh ^goserelin] 568
#72 ("ici 118630" or ici118630):ti,ab 7
#73 ("ZD-9393" or ZD9393):ti,ab 1
#74 zoladex:ti,ab 318
#75 leuprorelin:ti,ab 248
#76 carcinil:ti,ab 0
#77 enanton*:ti,ab 21
#78 ginecrin:ti,ab 1
#79 leuplin:ti,ab 7
#80 [mh ^Leuprolide] 686
#81 leuprolide:ti,ab 696
#82 lucrin:ti,ab 21
#83 lupron:ti,ab 77
#84 provren:ti,ab 0
#85 procrin:ti,ab 2
#86 ("tap 144" or tap144):ti,ab 24
#87 (a-43818 or a43818):ti,ab 0
#88 Trenantone:ti,ab 3
#89 staladex:ti,ab 0
#90 prostap:ti,ab 9
#91 [mh ^Nafarelin] 77
#92 nafarelin:ti,ab 114
#93 ("76932-56-4" or "76932564"):ti,ab 0
#94 ("76932-60-0" or "76932600"):ti,ab 2
#95 ("86220-42-0" or "86220420"):ti,ab 0
#96 ("rs 94991 298" or rs94991298):ti,ab 0
#97 synarel:ti,ab 10
#98 deslorelin:ti,ab 16

#99 gonadorelin:ti,ab 11
#100 ("33515-09-2" or "33515092"):ti,ab 0
#101 ("51952-41-1" or "51952411"):ti,ab 0
#102 ("52699-48-6" or "52699486"):ti,ab 0
#103 cetrorelix:ti,ab 221
#104 cetrotide:ti,ab 111
#105 ("NS 75A" or NS75A):ti,ab 0
#106 ("NS 75B" or NS75B):ti,ab 0
#107 ("SB 075" or SB075):ti,ab 0
#108 ("SB 75" or SB75):ti,ab 10
#109 gonadoliberin:ti,ab 5
#110 kryptocur:ti,ab 0
#111 cetrorelix:ti,ab 221
#112 cetrotide:ti,ab 111
#113 antagon:ti,ab 12
#114 ganirelix:ti,ab 142
#115 ("ORG 37462" or ORG37462):ti,ab 4
#116 orgalutran:ti,ab 45
#117 ("RS 26306" or RS26306):ti,ab 0
#118 ("AY 24031" or AY24031):ti,ab 0
#119 factrel:ti,ab 1
#120 fertagyl:ti,ab 0
#121 lutrelef:ti,ab 0
#122 lutrepulse:ti,ab 1
#123 relefact:ti,ab 1
#124 fertiral:ti,ab 0
#125 (hoe471 or "hoe 471"):ti,ab 3
#126 relisorm:ti,ab 0
#127 cystorelin:ti,ab 0
#128 dirigestran:ti,ab 0
#129 {or #33-#128} 6844
#130 #32 and #129 27
#131 #130 with Cochrane Library publication date Between Jan 2000 and Jul 2020, in Cochrane Reviews 1
#132 #130 27
#133 "conference":pt or (clinicaltrials or trialsearch):so 492465
#134 #132 not #133 9
#135 #134 with Publication Year from 2000 to 2020, in Trials 8

Database: HTA
Platform: CRD
Version: HTA
Search date: 23/7/2020
Number of results retrieved: 26
Search strategy:

1 MeSH DESCRIPTOR Gender Dysphoria EXPLODE ALL TREES 0
2 MeSH DESCRIPTOR Gender Identity EXPLODE ALL TREES 14

3 MeSH DESCRIPTOR Sexual and Gender Disorders EXPLODE ALL TREES 2

4 MeSH DESCRIPTOR Transsexualism EXPLODE ALL TREES 12
5 MeSH DESCRIPTOR Transgender Persons EXPLODE ALL TREES 3
6 MeSH DESCRIPTOR Health Services for Transgender Persons EXPLODE ALL TREES 0
7 MeSH DESCRIPTOR Sex Reassignment Procedures EXPLODE ALL TREES 1

8 ((gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*))) 28
9 ((transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*)) 76
10 ((trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*)) 83
11 (((sex or gender*) adj3 (reassign* or chang* or transform* or transition*))) 24
12 (male-to-female or m2f or female-to-male or f2m) 86
13 ((transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*)) 0
14 #1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13 262
15 (#1 OR #2 OR #3 OR #4 OR #5 OR #6 OR #7 OR #8 OR #9 OR #10 OR #11 OR #12 OR #13) IN HTA 30

*26 results are from 200 onwards. Downloaded as a set to sift for drug terms rather than continuing with search strategy.

Database: APA PsycInfo
Search date: July 2020 (Week 2)
Search Strategy:
---
1 Gender Dysphoria/ (936)
2 Gender Identity/ (8648)
3 Transsexualism/ (2825)
4 Transgender/ (5257)
5 exp Gender Reassignment/ (568)
6 (gender* adj3 (dysphori* or affirm* or incongruen* or identi* or disorder* or confus* or minorit* or queer*)).tw. (15471)
7 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. (13028)
8 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. (7679)
9 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).tw. (5796)
10 (male-to-female or m2f or female-to-male or f2m).tw. (63688)
11 or/1-10 (99560)
12 exp Infant Development/ (21841)
13 (prematur* or pre-matur* or preterm* or pre-term* or infan* or newborn* or new-born* or perinat* or peri-nat* or neonat* or neo-nat* or baby* or babies or toddler*).ti,ab,in,jn. (150219)

14 Child Characteristics/ or exp Child Behavior/ or Child Psychology/ or exp Child Welfare/ or Child Psychiatry/ (23423)
15 (child* or minor or minors or boy* or girl* or kid or kids or young*).ti,ab,in,jn. (984230)
16 (pediatric* or paediatric* or peadiatric*).ti,ab,in,jn. (78962)
17 Adolescent Psychiatry/ or Adolescent Behavior/ or Adolescent Development/ or Adolescent Psychology/ or Adolescent Characteristics/ or Adolescent Health/ (62142)
18 Puberty/ (2753)
19 (adolescen* or pubescen* or prepubescen* or pre-pubescen* or pubert* or prepubert* or pre-pubert* or teen* or preteen* or pre-teen* or juvenil* or youth* or under*age*).ti,ab,in,jn. (347604)
20 Schools/ or exp elementary school students/ or high school students/ or junior high school students/ or middle school students/ (113053)
21 Child Day Care/ or Nursery Schools/ (2836)
22 (pre-school* or preschool* or kindergar* or daycare or day-care or nurser* or school* or pupil* or student*).ti,ab,jn. (772814)
23 (("eight" or "nine" or "ten" or "eleven" or "twelve" or "thirteen" or "fourteen" or "fifteen" or "sixteen" or "seventeen" or "eighteen" or "nineteen") adj2 (year or years or age or ages or aged)).ti,ab. (21475)
24 (("8" or "9" or "10" or "11" or "12" or "13" or "14" or "15" or "16" or "17" or "18" or "19") adj2 (year or years or age or ages or aged)).ti,ab. (285697)
25 or/12-24 (1772959)
26 11 and 25 (49612)
27 (transchild* or transyouth* or transteen* or transadoles* or transgirl* or transboy*).tw. (14)
28 26 or 27 (49613)
29 exp Gonadotropic Hormones/ (4226)
30 (pubert* adj3 block*).ti,ab. (29)
31 ((gonadotrophin or gonadotropin) and releasing).ti,ab. (1060)
32 (GnRH adj2 analog*).ti,ab. (49)
33 GnRH*.ti,ab. (998)
34 "GnRH agonist*".ti,ab. (72)
35 triptorelin.ti,ab. (25)
36 arvekap.ti,ab. (0)
37 ("AY 25650" or AY25650).ti,ab. (0)
38 ("BIM 21003" or BIM21003).ti,ab. (0)
39 ("BN 52014" or BN52014).ti,ab. (0)
40 ("CL 118532" or CL118532).ti,ab. (0)
41 Debio.ti,ab. (7)
42 diphereline.ti,ab. (0)
43 moapar.ti,ab. (0)
44 pamorelin.ti,ab. (0)
45 trelstar.ti,ab. (0)
46 triptodur.ti,ab. (0)
47 ("WY 42422" or WY42422).ti,ab. (0)
48 ("WY 42462" or WY42462).ti,ab. (0)
49 gonapeptyl.ti,ab. (0)
50 decapeptyl.ti,ab. (3)
51 salvacyl.ti,ab. (1)

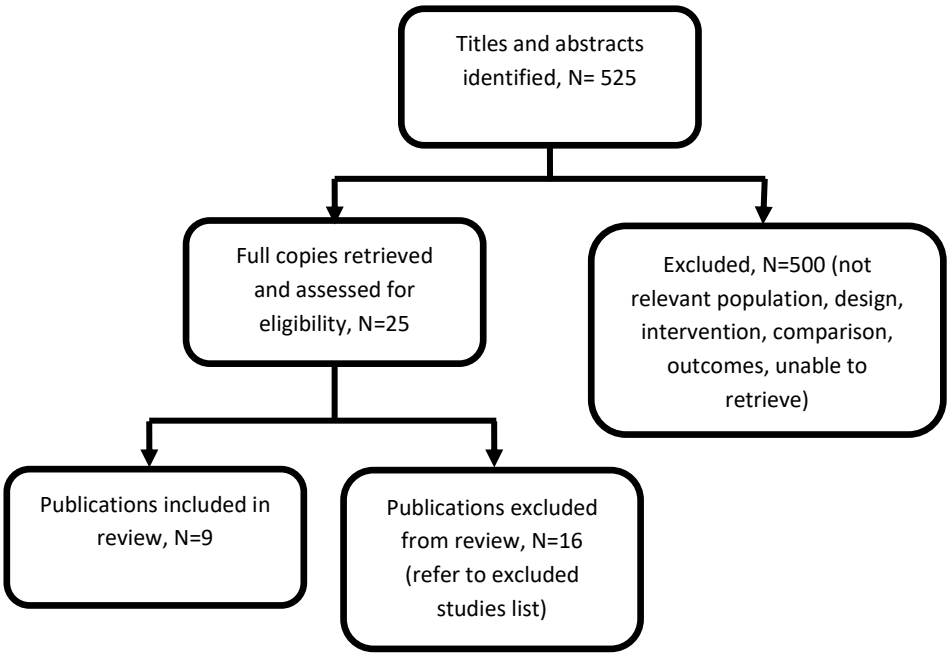
52 buserelin.ti,ab. (6)
53 bigonist.ti,ab. (0)
54 ("hoe 766" or hoe-766 or hoe766).ti,ab. (0)
55 profact.ti,ab. (0)
56 receptal.ti,ab. (0)
57 suprecur.ti,ab. (0)
58 suprefact.ti,ab. (0)
59 tiloryth.ti,ab. (0)
60 histrelin.ti,ab. (1)
61 "LHRH-hydrogel implant".ti,ab. (0)
62 ("RL 0903" or RL0903).ti,ab. (0)
63 ("SPD 424" or SPD424).ti,ab. (0)
64 goserelin.ti,ab. (30)
65 ("ici 118630" or ici118630).ti,ab. (0)
66 ("ZD-9393" or ZD9393).ti,ab. (0)
67 zoladex.ti,ab. (3)
68 leuprorelin.ti,ab. (12)
69 carcinil.ti,ab. (0)
70 enanton*.ti,ab. (1)
71 ginecrin.ti,ab. (0)
72 leuplin.ti,ab. (0)
73 leuprolide.ti,ab. (79)
74 lucrin.ti,ab. (1)
75 lupron.ti,ab. (18)
76 provren.ti,ab. (0)
77 procrin.ti,ab. (0)
78 ("tap 144" or tap144).ti,ab. (1)
79 (a-43818 or a43818).ti,ab. (0)
80 Trenantone.ti,ab. (0)
81 staladex.ti,ab. (0)
82 prostap.ti,ab. (0)
83 nafarelin.ti,ab. (1)
84 ("76932-56-4" or "76932564").ti,ab. (0)
85 ("76932-60-0" or "76932600").ti,ab. (0)
86 ("86220-42-0" or "86220420").ti,ab. (0)
87 ("rs 94991 298" or rs94991298).ti,ab. (0)
88 synarel.ti,ab. (0)
89 deslorelin.ti,ab. (8)
90 gonadorelin.ti,ab. (3)
91 ("33515-09-2" or "33515092").ti,ab. (0)
92 ("51952-41-1" or "51952411").ti,ab. (0)
93 ("52699-48-6" or "52699486").ti,ab. (0)
94 cetrorelix.ti,ab. (9)
95 cetrotide.ti,ab. (0)
96 ("NS 75A" or NS75A).ti,ab. (0)
97 ("NS 75B" or NS75B).ti,ab. (0)
98 ("SB 075" or SB075).ti,ab. (0)
99 ("SB 75" or SB75).ti,ab. (1)

100 gonadoliberin.ti,ab. (1)
101 kryptocur.ti,ab. (0)
102 cetrorelix.ti,ab. (9)
103 cetrotide.ti,ab. (0)
104 antagon.ti,ab. (0)
105 ganirelix.ti,ab. (0)
106 ("ORG 37462" or ORG37462).ti,ab. (0)
107 orgalutran.ti,ab. (0)
108 ("RS 26306" or RS26306).ti,ab. (0)
109 ("AY 24031" or AY24031).ti,ab. (0)
110 factrel.ti,ab. (0)
111 fertagyl.ti,ab. (0)
112 lutrelef.ti,ab. (0)
113 lutrepulse.ti,ab. (0)
114 relefact.ti,ab. (0)
115 fertiral.ti,ab. (0)
116 (hoe471 or "hoe 471").ti,ab. (0)
117 relisorm.ti,ab. (0)
118 cystorelin.ti,ab. (0)
119 dirigestran.ti,ab. (0)
120 or/29-119 (4869)
121 28 and 120 (130)
122 limit 121 to english language (120)
123 limit 122 to yr="2000 -Current" (93)

Appendix C Evidence selection

The literature searches identified 525 references. These were screened using their titles and abstracts and 25 references were obtained and assessed for relevance. Of these, 9 references are included in the evidence review. The remaining 16 references were excluded and are listed in appendix D.

Figure 1 – Study selection flow diagram



References submitted with Preliminary Policy Proposal

There is no preliminary policy proposal for this policy.

Appendix D Excluded studies table

Study reference	Reason for exclusion
Achille, C., Taggart, T., Eaton, N.R. et al. (2020) Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: Preliminary results. International Journal of Pediatric Endocrinology 2020(1): 8	Intervention – data for GnRH analogues not reported separately from other interventions
Bechard, Melanie, Vanderlaan, Doug P, Wood, Hayley et al. (2017) Psychosocial and Psychological Vulnerability in Adolescents with Gender Dysphoria: A "Proof of Principle" Study. Journal of sex & marital therapy 43(7): 678-688	Population – no GnRH analogues at time of study
Chew, Denise, Anderson, Jemma, Williams, Katrina et al. (2018) Hormonal Treatment in Young People With Gender Dysphoria: A Systematic Review. Pediatrics 141(4)	All primary studies included apart from 1 conference abstract
de Vries, Annelou L C, McGuire, Jenifer K et al. (2014) Young adult psychological outcome after puberty suppression and gender reassignment. Pediatrics 134(4): 696-704	Population – relevant population included in de Vries et al. 2011
Ghelani, Rahul, Lim, Cheryl, Brain, Caroline et al. (2020) Sudden sex hormone withdrawal and the effects on body composition in late pubertal adolescents with gender dysphoria. Journal of pediatric endocrinology & metabolism: JPEM 33(1): 107-112	Outcomes – not in the PICO

Study reference	Reason for exclusion
Giovanardi, G, Morales, P, Mirabella, M et al. (2019) Transition memories: experiences of trans adult women with hormone therapy and their beliefs on the usage of hormone blockers to suppress puberty. Journal of endocrinological investigation 42(10): 1231-1240	Population – adults only
Hewitt, Jacqueline K, Paul, Campbell, Kasiannan, Porpavai et al. (2012) Hormone treatment of gender identity disorder in a cohort of children and adolescents. The Medical journal of Australia 196(9): 578-81	Outcomes – no data reported for relevant outcomes
Jensen, R.K., Jensen, J.K., Simons, L.K. et al. (2019) Effect of Concurrent Gonadotropin-Releasing Hormone Agonist Treatment on Dose and Side Effects of Gender-Affirming Hormone Therapy in Adolescent Transgender Patients. Transgender Health 4(1): 300-303	Outcomes – not in the PICO
Klaver, Maartje, de Mutsert, Renee, Wiepjes, Chantal M et al. (2018) Early Hormonal Treatment Affects Body Composition and Body Shape in Young Transgender Adolescents. The journal of sexual medicine 15(2): 251-260	Outcomes – not in the PICO
Klaver, Maartje, de Mutsert, Renee van der Loos, Maria A T C et al. (2020) Hormonal Treatment and Cardiovascular Risk Profile in Transgender Adolescents. Pediatrics 145(3)	Outcomes – not in the PICO
Lopez, Carla Marisa, Solomon, Daniel, Boulware, Susan D et al. (2018) Trends in the use of puberty blockers among transgender children in the United States. Journal of pediatric endocrinology & metabolism : JPEM 31(6): 665-670	Outcomes – not in the PICO
Schagen, Sebastian E E, Lustenhouwer, Paul, Cohen-Kettenis, Peggy T et al. (2018) Changes in Adrenal Androgens During Puberty Suppression and Gender-Affirming Hormone Treatment in Adolescents With Gender Dysphoria. The journal of sexual medicine 15(9): 1357-1363	Outcomes – not in the PICO
Swendiman, Robert A, Vogiatzi, Maria G, Alter, Craig A et al. (2019) Histrelin implantation in the pediatric population: A 10-year institutional experience. Journal of pediatric surgery 54(7): 1457-1461	Population – less than 10% of participants had gender dysphoria; data not reported separately
Turban, Jack L, King, Dana, Carswell, Jeremi M et al. (2020) Pubertal Suppression for Transgender Youth and Risk of Suicidal Ideation. Pediatrics 145(2)	Intervention – data for GnRH analogues not reported separately from other interventions
Vrouenraets, Lieke Josephina Jeanne Johanna, Fredriks, A Miranda, Hannema, Sabine E et al. (2016) Perceptions of Sex, Gender, and Puberty Suppression: A Qualitative Analysis of Transgender Youth. Archives of sexual behavior 45(7): 1697-703	Outcomes – not in the PICO
Zucker, Kenneth J, Bradley, Susan J, Owen-Anderson, Allison et al. (2010) Puberty-blocking hormonal therapy for adolescents with gender identity disorder: A descriptive clinical study. Journal of Gay & Lesbian Mental Health 15(1): 58-82	Intervention – data for GnRH analogues not reported separately from other interventions

Appendix E Evidence tables

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Brik T, Vrouenraets L, de Vries M, et al. (2020) Trajectories of adolescents treated with gonadotropin-releasing hormone analogues for gender dysphoria. Archives of Sexual Behaviour https://doi.org/10.1007/s10508-020-01660-8 Netherlands Retrospective observational single-centre study To document trajectories after the initiation of GnRH analogue and explore reasons for extended use and discontinuation of GnRH analogues. Includes participants seen between November 2010 and January 1, 2018.	Inclusion criteria were adolescents with gender dysphoria, according to the DSM-5 criteria, seen at the single centre and treated with GnRH analogues between November 2010 and January 1, 2018. The study excluded adolescents without a diagnosis of gender dysphoria, those who had coexisting problems that interfered with the diagnostic process and/or might interfere with successful treatment (not further defined), those adolescents not wanting hormones, those with ongoing diagnostic evaluation and those who did not attend appointments. The sample consisted of 143 adolescents meeting the inclusion/exclusion criteria, 38 transfemales, 105 transmales, with median ages of 15.0 years (range 11.1 to 18.6 years) and 16.1 years	The study only reports that GnRH analogues were given, no specific drug, dose, route, or frequency of administration are reported. No comparator cohort was used in the study. Follow-up was at (up to) 9 years (last follow-up July 2019).	**Critical outcomes** No critical outcomes assessed. **Important outcomes** ***Psychosocial impact*** Not assessed. ***Engagement with health care services*** Not formally assessed but the study reported that out of 214 age and developmentally appropriate adolescents for potential inclusion in the study, 9 were excluded as they stopped attending appointments (4.2%). ***Stopping treatment*** Of the 143 adolescents, 9 (6.2%, 1 transfemale and 8 transmales) stopped taking GnRH analogues after a median duration of 0.8 years (range 0.1 to 3.0). Four adolescents (2.8%) discontinued GnRH analogues although they wanted to continue endocrine treatments for gender dysphoria: • 1 transmale stopped due to increase in mood problems, suicidal thoughts and confusion attributed to GnRH analogues (later had gender-affirming hormones at an adult gender clinic)[1] • 1 transmale experienced hot flushes, increased migraines, had a fear of injections, stress at school and unrelated medical issues, and	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection** 1. somewhat representative 2. no-non exposed cohort 3. secure record 4. yes **Domain 2: Comparability** 1. no comparator **Domain 3: Outcome** 1. record linkage 2. yes 3. complete follow-up **Overall quality is assessed as poor.** Other comments: Physical and psychological comorbidity was poorly reported, concomitant use of other medicines was not reported. Source of funding: not reported.

	(range 10.1 to 17.9 years), respectively at commencement of GnRH analogues. Of the 143 adolescents in the study, 125 (87%, 36 transfemales and 89 transmales) subsequently started treatment with gender-affirming hormones after median 1.0 (range 0.5 to 3.8) years and 0.8 (0.3 to 3.7) years, respectively. Median age at the start of gender-affirming hormones was 16.2 years (range 14.5 to 18.6 years) in transfemales and 17.1 years (range 14.9 to 18.8 years) in transmales. Five adolescents who used GnRH analogues had not started gender-affirming hormones at the time of data collection as they were not yet eligible for this treatment due to age. At the time of data collection, they had used GnRH analogues for a median duration of 2.1 years (range 1.6 to 2.8). Tanner stage was not reported. Six adolescents had been referred to a gender clinic elsewhere for further		temporarily discontinued treatment (after 4 months)[2] • 1 transmale experienced mood swings 4 months after commencing GnRH analogues. After 2.2 years he developed unexplained severe nausea and rapid weight loss and due to his general condition discontinued GnRH analogues after 2.4 years[3] • 1 transmale stopped GnRH analogues as his parents were unable to regularly collect medication from the pharmacy and take him to appointments for the injections[4] Five adolescents (3.5%) stopped treatment as they no longer wished to continue with gender-affirming treatment. • 1 adolescent had been very distressed about breast development at the start of GnRH analogues and later thought that she might want to live as a woman without breasts. She did not want to live as a boy and discontinued GnRH analogues, although dreaded breast development and menstruation. • 1 adolescent experienced concurrent psychosocial problems interfering with the exploration of gender identity and did not currently want treatment.[5] • 1 adolescent felt more in between male and female and therefore did not want to continue with GnRH analogues.[6] • 1 adolescent made a social transition while using GnRH	

	treatment, including 1 who had prolonged use.		analogues and shortly after decided to discontinue treatment.[7] • 1 adolescent discontinued after using GnRH analogues as the treatment allowed them to feel who they were.[8]	

[1] The adolescent later indicated "I was already fully matured when I started GnRH analogues, menstruations were already suppressed by contraceptives. For me, it had no added value" (transmale, age 19 years).
[2] The adolescent restarted endocrine treatment (testosterone) 5 months later.
[3] The adolescent recovered over the next 2 years and subsequently started lynestrenol and testosterone treatment.
[4] The adolescent subsequently started lynestrenol to suppress menses, he was not yet eligible for testosterone treatment.
[5] The adolescent later reflected that "The decision to stop GnRH analogues to my mind was made by the gender team, because they did not think gender dysphoria was the right diagnosis. I do still feel like a man, but for me it is okay to be just me instead of a he or a she, so for now I do not want any further treatment" (adolescent assigned female sex at birth, age 16 years).
[6] The adolescent stated "At the moment, I feel more like 'I am' instead of 'I am a woman' or 'I am a man'" (adolescent assigned female sex at birth, age 16 years).
[7] The adolescent stated that "he had fallen in love with a girl and had never had such feelings, which made him question his gender identity. At subsequent visits, he indicated that he was happy living as a man.
[8] The adolescent stated "After using GnRH analogues for the first time, I could feel who I was without the female hormones, this gave me peace of mind to think about my future. It was an inner feeling that said I am a woman" (adolescent assigned female sex at birth, age 18 years).

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Costa R, Dunsford M, Skagerberg E, et al. (2015) Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. Journal of Sexual Medicine 12(11):2206-14. United Kingdom Prospective longitudinal observational single centre cohort study Includes participants referred to the service between 2010 and 2014.	Adolescents with gender dysphoria who completed a 6-month diagnostic process using DSM-IV-TR criteria for gender dysphoria (comprising the gender dysphoria assessment and psychological interventions) either immediately eligible for treatment with GnRH analogues or delayed eligible for treatment with GnRH analogues (received psychological support without any physical intervention). No exclusion criteria were reported. The sample consisted of 201 adolescents (sex assigned at birth male to female ratio 1:1.6)	**Intervention** 101 individuals were assessed as being immediately eligible for use of GnRH analogues (no specific treatment, dose or route, or frequency of administration reported but all received psychological support). **Comparison** The analyses were between the immediately eligible	**Critical outcomes** ***Impact on gender dysphoria*** The Utrecht gender dysphoria scale (UGDS) was used to assess adolescents' gender dysphoria related discomfort. The Cronbach's alpha (α) for the study was reported as 0.76 to 0.88, suggesting good internal consistency. UGDS was only reported once, for 160 adolescents (50 sex assigned at birth males and 110 sex assigned at birth females). The assessment time point is not reported (baseline or follow-up) and the comparison for gender related discomfort was between sex assigned at birth males and sex assigned at birth females. Sex assigned at birth males had a mean (±SD) UGDS score of 51.6 [±9.7] versus sex assigned at birth	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection** 1. somewhat representative 2. drawn from the same community as the exposed cohort. 3. secure record 4. no **Domain 2: Comparability** 1. partial comparator **Domain 3: Outcome** 1. independent assessment (unclear if blinded) 2. yes 3. incomplete follow-up

	mean (±SD) age 15.52±1.41 years) from a sampling frame of 436 consecutive adolescents referred to the service between 2010 and 2014. The mean (±SD) age (n=201) at the start of GnRH analogues was 16.48 [±1.26], range 13 to 17 years. The interval from the start of the diagnostic procedure to the start of puberty suppression took approximately 1.5 years [±0.63] from baseline. None of the delayed eligible individuals received puberty suppression at the time of this study. Tanner stage was not reported.	and delayed eligible (n=100) adolescents, Baseline assessment (following diagnostic procedure) was followed by follow-up at 6 months from baseline (T1), 12 months from baseline (T2) and 18 months from baseline (T3).	females score of 56.1 [±4.3], *t*-test 4.07; p<0.001. ***Impact on mental health*** Not assessed. ***Impact on quality of life*** Not assessed. **Important outcomes** ***Psychosocial impact*** The Children's Global Assessment Scale (CGAS) was used to assess adolescents' psychosocial functioning. The CGAS was administered by psychologists, psychotherapists, and psychiatrists (intra-class correlation assessment was 0.76 ≤ Cronbach's α ≤0.94). At baseline, CGAS scores were not associated with any demographic variable, in both sex assigned at birth males and sex assigned at birth females (all p>0.1). In comparison with sex assigned at birth females, sex assigned at birth males had statistically significantly lower mean (±SD) baseline CGAS scores (55.4 [±12.7] versus 59.2 [11.8]; *t*-test 2.15; p=0.03). There was no statistically significant difference in mean (±SD) CGAS scores at baseline (T0) between immediately eligible adolescents and delayed eligible adolescents (n=201, 58.72 [±11.38] versus 56.63 [±13.14]; *t*-test 1.21; p=0.23). **Immediately eligible compared with delayed eligible participants** At follow-up, there was no statistically significant difference in mean (±SD)	**Overall quality is assessed as poor.** Other comments: Physical and psychological comorbidity was poorly reported, concomitant use of other medicines was not reported. Large unexplained loss to follow-up (64.7%) at T3. Source of funding: not reported.

			CGAS scores at any follow-up time point (T1, T2 or T3) between immediately eligible adolescents and delayed eligible adolescents: • T1, n=201, 60.89 [±12.17] versus 60.29 [±12.81]; *t*-test 0.34; p=0.73 • T2, n=121, 64.70 [±13.34] versus 62.97 [±14.10]; *t*-test 0.69; p=0.49 • T3, n=71, 67.40 [±13.93] versus 62.53 [±13.54]; *t*-test 1.49; p=0.14. **All participants** There was a statistically significant increase in mean (±SD) CGAS scores at any follow-up time point (T1, T2 or T3) compared with baseline (T0) for the all adolescents group: • T0 (n=201) versus T1 (n=201), 57.73 [±12.27] versus 60.68 [±12.47]; *t*-test 4.87; p<0.001 • T0 (n=201) versus T2 (n=121), 57.73 [±12.27] versus 63.31 [±14.41]; *t*-test 3.70; p<0.001 • T0 (n=201) versus T3 (n=71), 57.73 [±12.27] versus 64.93 [±13.85]; *t*-test 4.11; p<0.001 There was a statistically significant increase in mean (±SD) CGAS scores when comparing the follow-up period T1 to T3 but not for the periods T1 to T2 and T2 to T3, for all adolescents: • T1 (n=201) versus T2 (n=121), 60.68 [±12.47] versus 63.31 [±14.41]; *t*-test 1.73; p<0.08 • T1 (n=201) versus T3 (n=71), 60.68 [±12.47] versus 64.93 [±13.85], *t*-test 2.40; p<0.02 • T2 (n=121) versus T3 (n=71), 63.31 [±14.41] versus 64.93 [±13.85], *t*-test 0.76; p=0.45	

			There were no statistically significant differences in CGAS scores between sex assigned at birth males and sex assigned at birth females with gender dysphoria in all the follow-up evaluations (all p>0.1). Delayed eligible and immediately eligible adolescents with gender dysphoria were not statistically significantly different for demographic variables (all p>0.1). **Immediately eligible participants** There was a statistically significant increase in mean (±SD) CGAS scores at follow-up times T2 and T3 compared with baseline (T0) but not for T0 versus T1, for the immediately eligible adolescents: • T0 (n=101) versus T1 (n=101), 58.72 [±11.38] versus 60.89 [±12.17]; *t*-test 1.31; p=0.19 • T0 (n=101) versus T2 (n=60), 58.72 [±11.38] versus 64.70 [±13.34]; *t*-test 3.02; p=0.003 • T0 (n=101) versus T3 (n=35), 58.72 [±11.38] versus 67.40 [±13.93]; *t*-test 3.66; p<0.001 There was a statistically significant increase in mean (±SD) CGAS scores when comparing the follow-up period T1 to T3 with each other but not for the periods T1 to T2 and T2 to T3, for the immediately eligible adolescents: • T1 (n=101) versus T2 (n=60), 60.89 [±12.17] versus 64.70 [±13.34]; *t*-test 1.85; p=0.07 • T1 (n=101) versus T3 (n=35), 60.89 [±12.17] versus 67.40 [±13.93], *t*-test 2.63; p<0.001	

			• T2 (n=60) versus T3 (n=35), 64.70 [±13.34] versus 67.40 [±13.93], *t*-test 0.94; p=0.35 The immediately eligible adolescents had a CGAS score which was not statistically significantly different compared to the sample of children/ adolescents without observed psychological /psychiatric symptoms after 12 months of puberty suppression (T3, *t*=0.01, *p*=0.99).	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
de Vries A, Steensma T, Doreleijers T, et al. (2011) Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. The Journal of Sexual Medicine 8 (8):2276-83. Netherlands Prospective longitudinal observational single centre before and after study.	The sample size was 70 adolescents receiving GnRH analogues (mean age [±SD] at assessment 13.6±1.8 years) from a sampling frame of 196 consecutive adolescents referred to the service between 2000 and 2008. Inclusion criteria were if they subsequently started gender-affirming hormones between 2003 and 2009 (mean [±SD] age at start of GnRH analogues was 14.75 [±1.92] years)[1]. No specific exclusion criteria were described. No diagnostic criteria or concomitant treatments were reported. Tanner stage of the included adolescents was not reported.	**Intervention** 70 adolescents were assessed at baseline (T0) before the start of GnRH analogues (no specific treatment, dose or route of administration reported). **Comparison** The same 70 adolescents were assessed again at follow-up (T1), shortly before starting gender-affirming hormones. Not all adolescents completed all assessments for all items[2].	**Critical outcomes** ***Impact on gender dysphoria*** Impact on gender dysphoria was assessed using the Utrecht Gender Dysphoria Scale (UGDS). • There was no statistically significant difference in UGDS scores between T0 and T1 (n=41). There was a statistically significant difference between sex assigned at birth males and sex assigned at birth females, with sex assigned at birth females reporting more gender dysphoria, *F* (*df, errdf*), *P*: 15.98 (1,39), p<0.001. ***Impact on mental health*** Depressive symptoms were assessed using the Beck Depression Inventory (BDI-II). • There was a statistically significant reduction in BDI score between T0 and T1, n=41, 8.31 [±7.12] versus 4.95 [±6.72], *F* (*df, errdf*), *P*: 9.28 (1,39), p=0.004. • There was no statistically significant difference between sex assigned at	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection** 1. somewhat representative of children and adolescents who have gender dysphoria 2. no non-exposed cohort 3. no description 4. no **Domain 2: Comparability** 1. study controls for age, age at start of treatment, IQ, and parental factors **Domain 3: Outcome** 1. no description 2. no/unclear 3. complete **Overall quality is assessed as poor**. Other comments: Physical and psychological comorbidity was not reported, concomitant use of

			birth males and sex assigned at birth females, *F* (*df, errdf*), *P*: 3.85 (1,39), p=0.057. Anger and anxiety were assessed using Trait Anger and Anxiety (TPI and STAI, respectively) Scales of the State-Trait Personality Inventory. • There was no statistically significant difference in anger (TPI) scale scores between T0 and T1 (n=41). There was a statistically significant difference between sex assigned at birth males and sex assigned at birth females, with sex assigned at birth females reporting increased anger compared with sex assigned at birth males, *F* (*df, errdf*), *P*: 5.70 (1,39), p=0.022. • Similarly, there was no statistically significant difference in anxiety (STAI) scale scores between T0 and T1 (n=41). There was a statistically significant difference between sex assigned at birth males and sex assigned at birth females, with sex assigned at birth females reporting increased anxiety compared with sex assigned at birth males, *F* (*df, errdf*), *P*: 16.07 (1,39), p<0.001. ***Impact on quality of life*** Not assessed. **Important outcomes** ***Impact on body image*** Impact on body image was assessed using the Body Image Scale to measure body satisfaction (BIS).	other medicines was not reported. Source of funding: This study was supported by a personal grant awarded to the first author by the Netherlands Organization for Health Research and Development.

			There was no statistically significant difference between T0 and T1 for any of the 3 BIS scores (primary sex characteristics, secondary sex characteristics or neutral characteristics, n=57). There were statistically significant differences between sex assigned at birth males and sex assigned at birth females, with sex assigned at birth females reporting more dissatisfaction, for: • primary sexual characteristics, *F* (*df, errdf*), *P*: 4.11 (1,55), p=0.047. • secondary sexual characteristics, *F* (*df, errdf*), *P*: 11.57 (1,55), p=0.001. But no statistically significant difference between sex assigned at birth males and sex assigned at birth females was found for neutral characteristics. However, there was a significant interaction effect between sex assigned at birth sex and the changes of gender dysphoria between T0 and T1; sex assigned at birth females became more dissatisfied with their secondary sex characteristics compared with sex assigned at birth males, *F* (*df, errdf*), *P*: 14.59 (1,55), p<0.001) and neutral characteristics, *F* (*df, errdf*), *P*: 15.26 (1,55), p<0.001). ***Psychosocial impact*** Psychosocial impact was assessed using both the Child Behaviour Checklist (CBCL) and the Youth Self-Report (YSR) to parents and adolescents, respectively. The Children's Global Assessment Scale was also reported. There was a statistically significant decrease in mean (±SD) total, internalising, and externalising[3] parental	

			CBCL scores between T0 and T1[4] for all adolescents (n=54): • Total score (T0 – T1) 60.70 [±12.76] versus 54.46 [±11.23], *F* (*df, errdf*), *P*: 26.17 (1,52), p<0.001. • Internalising score (T0 – T1) 61.00 [±12.21] versus 54.56 [±10.22], *F* (*df, errdf*), *P*: 22.93 (1,52), p<0.001. • Externalising score (T0 – T1) 58.04 [±12.99] versus 53.81 [±11.86], *F* (*df, errdf*), *P*: 12.04 (1,52), p=0.001. There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for total and internalising CBCL score but there was a significant difference for the externalising score: • Externalising score, *F* (*df, errdf*), *P*: 6.29 (1,52), p=0.015. There was a statistically significant decrease in mean (±SD) total, internalising, and externalising[3] YSR scores between T0 and T1 for all adolescents (n=54): • Total score (T0 – T1) 55.46 [±11.56] versus 50.00 [±10.56], *F* (*df, errdf*), *P*: 16.24 (1,52), p<0.001. • Internalising score (T0 – T1) 56.04 [±12.49] versus 49.78 [±11.63], *F* (*df, errdf*), *P*: 15.05 (1,52), p<0.001. • Externalising score (T0 – T1) 53.30 [±11.87] versus 49.98 [±9.35], *F* (*df, errdf*), *P*: 7.26 (1,52), p=0.009. There was no statistically significant difference between sex assigned at birth males and sex assigned at birth females for total and internalising YSR score but there was a significant difference for the externalising score:	

			• Externalising score, *F* (*df, errdf*), *P*: 9.14 (1,52), p=0.004. There was a statistically significant increase in CGAS mean (±SD) score between T0 and T1 (n=41), 70.24 [±10.12] versus 73.90 [±9.63], *F* (*df, errdf*), *P*: 8.76 (1,39), p=0.005. There was a statistically significant difference between sex assigned at birth males and sex assigned at birth females, with sex assigned at birth females reporting lower score for global functioning compared with sex assigned at birth males, *F* (*df, errdf*), *P*: 5.77 (1,52), p=0.021. The proportion of adolescents scoring in the clinical range significantly decreased between T0 and T1, on the CBCL total problem scale (44.4% versus 22.2%, X^2[1] = 6.00, p=0.001), and the internalising scale (29.6% versus 11.1%, X^2[1] = 5.71, p=0.017) of the YSR.	

[1] There were statistically significant mean age [±SD] differences between sex assigned at birth males and sex assigned at birth females for age at assessment (13.14 [±1.55] versus 14.10 [±1.99] years, p=0.028), age at start of GnRH analogues (14.25 [±1.79] versus 15.21 [±1.95] years, p=0.036) and age at the start of gender-affirming hormones (16.24 [±1.21] versus 16.99 [±1.09] years, p=0.008). No statistically significant differences were seen for other baseline characteristics, time between GnRH analogue and gender-affirming hormones, full scale IQ, parental marital status, education, and sexual attraction to own, other or both sexes.
[2] Independent t-tests between mean scores on the CBCL, YSR, BDI, TPI, STAI, CGAS, UGS, and BIS of adolescents who completed both assessments and mean scores of adolescents who completed only one of the assessments revealed no significant differences on all used measures, at neither T0 or at T1.
[3] The CBCL/YSR has 2 components: Internalising score which sums the anxious/depressed, withdrawn-depressed, and somatic complaints scores; externalising score which sums rule-breaking and aggressive behaviour. The total problems score is the sum of the scores of all the problem items. The YSR is a child self-report version of the CBCL.
[4] A repeated measures ANOVA (analysis of variance) was used.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Joseph T, Ting J, Butler G. (2019) The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort. Journal of pediatric endocrinology & metabolism 32(10): 1077-1081	Adolescents (12 to 14 years) with gender dysphoria (no diagnostic criteria described), n=70, including 31 transfemales and 39 transmales.	Treatment with a GnRH analogue for at least 1 year or ongoing until they reached 16 years. No specific treatment, dose or route of	**Critical outcomes** No critical outcomes assessed. **Important outcomes** ***Bone density: lumbar[1]*** **Lumbar spine bone mineral apparent density (BMAD)[2] 0 to 1 year** Transfemales (mean [±SD]):	This study was appraised using the Newcastle-Ottawa quality assessment checklist for cohort studies. **Domain 1: Selection**

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
United Kingdom Retrospective longitudinal observational single centre study To investigate whether there is any significant loss of bone mineral density (BMD) and bone mineral apparent density (BMAD) for up to 3 years of GnRH analogues. To investigate whether there was a significant drop after 1 year of treatment following abrupt withdrawal. 2011 to 2016	All had been seen and assessed by a Gender Identity Development Service multi-disciplinary psychosocial health team for at least 4 assessments over a minimum of 6 months. All participants had entered puberty and all but 2 of the transmales were postmenarchal. 57% of the transfemales were in early puberty (G2–3 and testicular volume >4 mL) and 43% were in late puberty (G4–5). Details of the sampling frame were not reported. Further details of how the sample was drawn are not reported.	administration reported. No concomitant treatments were reported. No comparator.	0.235 (0.030) g/cm3 at baseline, 0.233 g/cm3 (0.029) at 1 year (p=0.459); z-score 0.859 (0.154) at baseline, −0.228 (1.027) at 1 year (p=0.000) Transmales (mean [±SD]): 0.196 (0.035) g/cm3 at baseline, 0.201 (0.033) g/cm3 at 1 year (p=0.074); z-score −0.186 (1.230) at baseline, −0.541 (1.396) at 1 year (p=0.006) **Lumbar spine BMAD 0 to 2 years** Transfemales (mean [±SD]): 0.240 (0.027) g/cm3 at baseline, 0.240 (0.030) g/cm3 at 2 years (p=0.865); z-score 0.486 (0.809) at baseline, −0.279 (0.930) at 2 years (p=0.000) Transmales (mean [±SD]): 0.195 (0.058) g/cm3 at baseline, 0.198 (0.055) at 2 years (p=0.433); z-score −0.361 (1.439) at baseline, −0.913 (1.318) at 2 years (p=0.001) **Lumbar spine bone mineral density (BMD) 0 to 1 year** Transfemales (mean [±SD]): 0.860 (0.154) kg/m2 at baseline, 0.859 (0.129) kg/m2 at 1 year (p=0.962); z-score −0.016 (1.106) at baseline, −0.461 (1.121) at 1 year (p=0.003) Transmales (mean [±SD]): 0.694 (0.149) kg/m2 at baseline, 0.718 (0.124) kg/m2 at 1 year (p=0.006); z-score −0.395 (1.428) at baseline, −1.276 (1.410) at 1 year (p=0.000) **Lumbar spine BMD 0 to 2 years** Transfemales (mean [±SD]): 0.867 (0.141) kg/m2 at baseline, 0.878 (0.130) kg/m2 at 2 years (p=0.395); z-score 0.130 (0.972) at baseline, −0.890 (1.075) at 2 years (p=0.000) Transmales (mean [±SD]):	1. Somewhat representative of children and adolescents who have gender dysphoria 2. Not applicable 3. Via routine clinical records 4. No **Domain 2: Comparability** 1. No control group **Domain 3: Outcome** 1. Via routine clinical records 2. Yes 3. No statement **Overall quality is assessed as poor**. Other comments: although the evidence is of poor quality, the results suggest a possible association between GnRH analogues and BMAD. However, the results are not reliable and could be due to bias or chance. Further details of how the sample was drawn are not reported. No concomitant treatments were reported. Source of funding: None disclosed

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			0.695 (0.220) kg/m2 at baseline, 0.731 (0.209) kg/m2 at 2 years (p=0.058); z-score −0.715 (1.406) at baseline, −2.000 (1.384) at 2 years (p=0.000) ***Bone density: femoral*** **Femoral neck (hip) BMD 0 to 1 year** Transfemales (mean [±SD]): 0.894 (0.118) kg/m2 at baseline, 0.905 (0.104) kg/m2 at 1 year (p=0.571); z-score 0.157 (0.905) at baseline, −0.340 (0.816) at 1 year (p=0.002) Transmales (mean [±SD]): 0.772 (0.137) kg/m2 at baseline, 0.785 (0.120) kg/m2 at 1 year (p=0.797); z-score −0.863 (1.215) at baseline, −1.440 (1.075) at 1 year (p=0.000) **Femoral neck (hip) BMD 0 to 2 years** Transfemales (mean [±SD]): 0.920 (0.116) kg/m2 at baseline, 0.910 (0.125) kg/m2 at 2 years (p=0.402); z-score 0.450 (0.781) at baseline, −0.600 (1.059) at 2 years (p=0.002) Transmales (mean [±SD]): 0.766 (0.215) kg/m2 at baseline, 0.773 (0.197) at 2 years (p=0.604); z-score −1.075 (1.145) at baseline, −1.779 (0.816) at 2 years (p=0.001)	

[1] Lumbar spine (L1-L4) BMD was measured by yearly dual energy X-ray absorptiometry (DXA) scans at baseline (n=70), 1 year (n=70), and 2 years (n=31).
[2] BMAD is a size adjusted value of BMD incorporating body size measurements using UK norms in growing adolescents. Reported as g/cm3 and z-scores. Hip BMAD z-scores were not calculated as there were no available reference ranges.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Khatchadourian K, Shazhan A, Metzger D. (2014) Clinical management of youth with gender dysphoria in	27 young people with gender dysphoria who started GnRH analogues (at mean age [±SD] 14.7±1.9 years) out of 84 young	**Intervention** 84 young people with gender dysphoria were included. For GnRH analogues no	**Critical Outcomes** No critical outcomes assessed. **Important outcomes** ***Stopping treatment***	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection**

Vancouver. The Journal of Pediatrics 164 (4): 906-11.
 	
Canada
 	
Retrospective observational chart review single centre study
	people seen at the unit between 1998 and 2011.
	Note: the transmale and transfemale subgroups reported in the paper is discrepant, 15 transmales and 11 transfemales (n=26) reported in the outcomes section rather than the n=27 stated in the paper; complete outcome reporting is also incomplete for the transfemale group.
	Inclusion criteria were at least Tanner stage 2 pubertal development, previous assessment by a mental health professional and a confirmed diagnosis of gender dysphoria (diagnostic criteria not specified). No exclusion criteria are specified.
		specific treatment, dose or route of administration reported.
		Comparison
		No comparator.
			The authors report that of 15 transmales taking GnRH analogues:
			• 14 transitioned to testosterone treatment during the observation period
			• 7 continued taking GnRH analogues after starting testosterone
			• 7 discontinued GnRH analogues after a median of 3.0 years (range 0.2 to 9.2 years), of which:
			  ○ 5 discontinued after hysterectomy and salpingo-oophorectomy
			  ○ 1 discontinued after 2.2 years (transitioned to gender-affirming hormone)
			  ○ 1 discontinued after <2 months due to mood and emotional lability
			The authors report that of 11 transfemales taking GnRH analogues:
			• 5 received oestrogen treatment during the observation period
			• 4 continued taking GnRH analogues during oestrogen treatment
			• 1 discontinued GnRH analogues during oestrogen treatment (no reason reported)
			• 1 stopped GnRH analogues after a few months due to emotional lability
			• 1 stopped GnRH analogues before oestrogen treatment (the following year delayed due to heavy smoking)
			• 1 discontinued GnRH analogues after 13 months due to choosing not to pursue transition
			
			Safety
			Of the 27 patients treated with GnRH analogues:
				1. not reported
				2. no non-exposed cohort
				3. secure record
				4. no
				Domain 2: Comparability
				1. not applicable
				Domain 3: Outcome
				1. record linkage
				2. yes
				3. in complete missing data
				
				Overall quality is assessed as poor.
				
				Other comments: mental health comorbidity was reported for all participants but not for the GnRH analogue cohort separately. Concomitant use of other medicines was not reported.
				
				Source of funding: No source of funding identified.

			• 1 transmale participant developed sterile abscesses; they were switched from leuprolide acetate to triptorelin, and this was well tolerated. • 1 transmale participant developed leg pains and headaches on GnRH analogues, which eventually resolved without treatment. • 1 participant gained 19 kg within 9 months of initiating GnRH analogues, although their body mass index was >85 percentile before GnRH analogues.	

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Klink D, Caris M, Heijboer A et al. (2015) Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. The Journal of clinical endocrinology and metabolism 100(2): e270-5 Netherlands Retrospective longitudinal observational single centre study To assess BMD development during GnRH analogues and at age 22 years in adolescents with gender dysphoria who started treatment for gender dysphoria during adolescence.	34 adolescents (mean age ±SD 14.9±1.9 for transfemales and 15.0±2.0 for transmales at start of GnRH analogues). Participants were included if they met DSM-IV-TR criteria for gender identity disorder of adolescence and had been treated with GnRH analogues and gender-affirming hormones during their pubertal years. No concomitant treatments were reported.	The intervention was GnRH analogue monotherapy (triptorelin pamoate 3.75 mg subcutaneously every 4 weeks) followed by gender-affirming hormones from 16 years with discontinuation of GnRH analogue after gonadectomy. Median duration of GnRH analogue monotherapy in transfemales was 1.3 years (range, 0.5 to 3.8 years), and in transmales was 1.5 years	**Critical outcomes** No critical outcomes assessed. **Important outcomes** ***Bone density: lumbar*** **Lumbar spine bone mineral apparent density (BMAD)[1]** Change from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales (mean [±SD]): GnRH analogue: 0.22 (0.03) g/cm3, gender-affirming hormones: 0.22 (0.02) g/cm3 (NS); z-score GnRH analogue: −0.44 (1.10), gender-affirming hormones: −0.90 (0.80) (p=NS) Change from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales (mean [±SD]: GnRH analogue: 0.25 (0.03) g/cm3, gender-affirming hormones: 0.24 (0.02) g/cm3 (NS);	This study was appraised using the Newcastle-Ottawa quality assessment checklist for cohort studies. **Domain 1: Selection** 1. somewhat representative of children and adolescents who have gender dysphoria 2. not applicable 3. via routine clinical records 4. no **Domain 2: Comparability** 1. no control group **Domain 3: Outcome** 1. via routine clinical records 2. yes 3. follow-up rate variable across timepoints and no description of those lost **Overall quality is assessed as poor**.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
1998 to 2012		(range, 0.25 to 5.2 years).	z-score GnRH analogue: 0.28 (0.90), gender-affirming hormones: −0.50 (0.81) (p=0.004) **Lumbar spine bone mineral density (BMD)[1]** Change from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales (mean [±SD]): GnRH analogue: 0.84 (0.13) g/m2, gender-affirming hormones: 0.84 (0.11) g/m2 (NS); z-score GnRH analogue: −0.77 (0.89), gender-affirming hormones: −1.01 (0.98) (NS) Change from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales (mean [±SD]): GnRH analogue: 0.95 (0.12) g/m2, gender-affirming hormones: 0.91 (0.10) g/m2 (p=0.006); z-score GnRH analogue: 0.17 (1.18), gender-affirming hormones: −0.72 (0.99) (p<0.001) ***Bone density; femoral*** **Femoral area BMAD[1]** Change from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales (mean [±SD]), GnRH analogue: 0.28 (0.04) g/cm3, gender-affirming hormones: 0.26 (0.04) g/cm3 (NS); z-score GnRH analogue: −0.93 (1.22), gender-affirming hormones: −1.57 (1.74) (p=NS) Change from starting GnRH analogue	Other comments: Within person comparison. Small numbers of participants in each subgroup. No concomitant treatments or comorbidities were reported. Source of funding: None disclosed

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			(mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales (mean [±SD]), GnRH analogue: 0.32 (0.04) g/cm3, gender-affirming hormones: 0.31 (0.04) (NS); z-score GnRH analogue: 0.01 (0.70), gender-affirming hormones: −0.28 (0.74) (NS) **Femoral area BMD[1]** Change from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales (mean [±SD]), GnRH analogue: 0.88 (0.12) g/m2, gender-affirming hormones: 0.87 (0.08) (NS); z-score GnRH analogue: −0.66 (0.77), gender-affirming hormones: −0.95 (0.63) (NS) Change from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales (mean [±SD]), GnRH analogue: 0.92 (0.10) g/m2, gender-affirming hormones: 0.88 (0.09) (p=0.005); z-score GnRH analogue: 0.36 (0.88), gender-affirming hormones: −0.35 (0.79) (p=0.001)	

[1] BMD and BMAD of the lumbar spine and femoral region (nondominant side) measured by DXA scans at start of GnRH analogues, (n=32), start of gender-affirming hormones (n=34), and at 22 years (n=34).

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Schagen SEE, Cohen-Kettenis PT, Delemarre-van de Waal HA et al. (2016)	Adolescents with gender dysphoria (n=116), median age (range) 13.6 years (11.6 to 17.9) in transfemales and 14.2 years (11.1 to	GnRH analogue monotherapy (triptorelin pamoate 3.75 mg at 0, 2 and 4	**Critical outcomes** No critical outcomes assessed. **Important outcomes**	This study was appraised using the Newcastle-Ottawa quality assessment checklist for cohort studies.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Efficacy and Safety of Gonadotropin-Releasing Hormone Agonist Treatment to Suppress Puberty in Gender Dysphoric Adolescents. The journal of sexual medicine 13(7): 1125-32 Netherlands Prospective longitudinal study To describe the changes in Tanner stage, testicular volume, gonadotropins, and sex steroids during GnRH analogues of adolescents with gender dysphoria to evaluate the efficacy. To report on liver enzymes, renal function and changes in body composition. 1998 to 2009	18.6) in transmales during first year of GnRH analogues. Participants were included if they met DSM-IV-TR criteria for gender dysphoria, had lifelong extreme gender dysphoria, were psychologically stable and were living in a supportive environment. No concomitant treatments were reported.	weeks followed by injections every 4 weeks, route of administration not described) for at least 3 months.	***Other safety outcomes: liver function*** Glutamyl transferase was not elevated at baseline or during treatment in any subject. Mild elevations of aspartate aminotransferase (AST) and alanine aminotransferase (ALT) above the reference range were present at baseline but were not more prevalent during treatment than at baseline. Glutamyl transferase, AST, and ALT levels did not significantly change from baseline to 12 months of treatment. No values or statistical analyses were reported. ***Other safety outcomes: kidney function*** **Change in serum creatinine between 0 and 1 year** Transfemales (mean [±SD]): 70 (12) micromol/l at baseline, 66 (13) micromol/l at 1 year (p=0.20) Transmales (mean [±SD]): 73 (8) micromol/l at baseline, 68 (13) micromol/l at 1 year (p=0.01)	**Domain 1: Selection** 1. somewhat representative of children and adolescents who have gender dysphoria 2. not applicable 3. via routine clinical records 4. no **Domain 2: Comparability** 1. no control group **Domain 3: Outcome** 1. via routine clinical records 2. yes 3. no statement **Overall quality is assessed as poor**. Other comments: Within person comparison. No concomitant treatments or comorbidities were reported. Source of funding: Ferring pharmaceuticals (triptorelin manufacturer)

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Staphorsius A, Baudewijntje P, Kreukels P, et al. (2015) Puberty suppression and executive functioning: an fMRI-study	The inclusion criteria were diagnosed with Gender Identity Disorder according to the DSM-IV-TR and at least 12 years old and Tanner stage of at least B2 or G2 to G3 with	**Intervention** GnRH analogues (triptorelin pamoate 3.75 mg every 4 weeks	**Critical Outcomes** No critical outcomes assessed. **Important outcomes** ***Psychosocial impact***	This study was appraised using the Newcastle-Ottawa tool for cohort studies. **Domain 1: Selection domain**

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
in adolescents with gender dysphoria. Psychoneuroendocrinology 565:190-9. Netherlands Cross-sectional (single time point) assessment single centre study	measurable oestradiol and testosterone levels in girls and boys, respectively. For all group's exclusion criteria were an insufficient command of the Dutch language (how assessed not reported), unadjusted endocrine disorders, neurological or psychiatric disorders that could lead to deviant test results (details not reported) use of psychotropic medication, and contraindications for an MRI scan. Additionally, adolescents receiving puberty delaying medication or any form of hormones besides oral contraceptives were excluded as controls. The sample size was 85 of whom 41 were adolescents (the numbers are discrepant with the number for whom outcomes are reported n=40) with gender dysphoria (20 of whom were being treated with GnRH analogues); 24 girls and 21 boys without gender dysphoria acted as controls (not further reported here). Details of the sampling frame are not reported. The ages at which GnRH analogues were started was not reported. The mean duration of treatment was 1.6 years (SD 1.0) Mean (±SD) Tanner stage for each group was reported: • Transfemales 3.9 [±1.1] • Transfemales on GnRH analogues 4.1 [±1.0]	subcutaneously or intramuscularly). **Comparison** The comparison was between adolescents with gender dysphoria receiving GnRH analogues and those without GnRH analogues.	The Child Behaviour Checklist (CBCL) was used to assess psychosocial impact. The CBCL was administered once during the study. The reported outcomes for each group were (n, mean [±SD]): • Transfemales (all, n=18) 57.8 [±9.2] • Transfemales on GnRH analogues (n=8) 57.4 [±9.8] • Transfemales without GnRH analogues (n=10) 58.2 [±9.3] • Transmales (all, n=22) 60.4 [±10.2] • Transmales on GnRH analogues (n=12) 57.5 [±9.4] • Transmales without GnRH analogues (n=10) 63.9 [±10.5] The analysis of the CBCL data is not discussed, and statistical analysis is unclear. ***Cognitive development or functioning*** **IQ[1]** • Transfemales (mean [±SD]) on GnRH analogues: 94.0 (10.3) • Transfemales (mean [±SD]) without GnRH analogues: 109.4 (21.2) • Transmales (mean [±SD]) on GnRH analogues: 95.8 (15.6) • Transmales (mean [±SD]) without GnRH analogues: 98.5 (15.9) **Reaction time[2]** • Transfemales (mean [±SD]) on GnRH analogues: 10.9 (4.1) • Transfemales (mean [±SD]) without GnRH analogues: 9.9 (3.1)	1. somewhat representative of children and adolescents who have gender dysphoria 2. drawn from the same community as the exposed cohort 3. via routine clinical records 4. no **Domain 2: Comparability** 1. study controls for age and diagnosis **Domain 3: Outcome** 1. via clinical assessment 2. yes 3. unclear **Overall quality is assessed as poor.** Other comments: Physical and psychological comorbidity was not reported, concomitant use of other medicines was not reported. Source of funding: This work was supported by an educational grant from the pharmaceutical firm Ferring BV, and by a VICI grant (453-08-003) from the Dutch Science Foundation. The authors state that funding sources did not play a role in any component of this study.

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
	• Transfemales without GnRH analogues 3.8 [±1.1] • Transmales 4.5 [±0.9] • Transmales on GnRH analogues 4.1 [±1.1] Transmales without GnRH analogues 4.9 [±0.3]		• Transmales (mean [±SD]) on GnRH analogues: 9.9 (3.1) • Transmales (mean [±SD]) without GnRH analogues: 10.0 (2.0) **Accuracy**[3] • Transfemales (mean [±SD]) on GnRH analogues: 73.9 (9.1) • Transfemales (mean [±SD]) without GnRH analogues: 83.4 (9.5) • Transmales (mean [±SD]) on GnRH analogues: 85.7 (10.5) • Transmales (mean [±SD]) without GnRH analogues: 88.8 (9.7)	

[1] Estimated with 4 subscales (arithmetic, vocabulary, picture arrangement, and block design) of the Wechsler Intelligence Scale for Children, third edition (WISC-III®, Wechsler 1991) or the Wechsler Adult Intelligence Scale, third edition (WAIS-III®, Wechsler 1997), depending on the participant's age.
[2] Reaction time in seconds in the Tower of London task
[3] Percentage of correct trials in the Tower of London task

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
Vlot, Mariska C, Klink, Daniel T, den Heijer, Martin et al. (2017) Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone 95: 11-19 Netherlands Retrospective observational data analysis study	Adolescents with gender dysphoria, n=70. Median age (range) 15.1 years (11.7 to 18.6) for transmales and 13.5 years (11.5 to 18.3) for transfemales at start of GnRH analogues. Participants were included if they had a diagnosis of gender dysphoria according to DSM-IV-TR criteria who were treated with GnRH analogues and then gender-affirming hormones. No concomitant treatments were reported. The study categorised	GnRH analogues (triptorelin pamoate 3.75 mg every 4 weeks subcutaneously).	**Critical outcomes** No critical outcomes reported **Important outcomes** ***Bone density: lumbar*** **Lumbar spine bone mineral apparent density (BMAD)** Change from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of <15 years; median [range]), GnRH analogue: 0.21 (0.17 to 0.25) g/cm3, gender-affirming hormones: 0.20 (0.18 to 0.24) g/cm3 (NS); z-score GnRH analogue: −0.20 (−1.82 to 1.18), gender-affirming hormones: −1.52 (−2.36 to 0.42) (p=0.001)	This study was appraised using the Newcastle-Ottawa quality assessment checklist for cohort studies. **Domain 1: Selection** 1. Somewhat representative of children and adolescents who have gender dysphoria 2. Not applicable 3. Via routine clinical records 4. No **Domain 2: Comparability** 1. No control group **Domain 3: Outcome** 1. Via routine clinical records 2. Yes

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
To investigate the course of 3 bone turnover markers in relation to bonemineral density, in adolescents with gender dysphoria during GnRH analogue and gender-affirming hormones. 2001 to 2011	participants into a young and old pubertal group, based on their bone age. The young transmales had a bone age of <14 years and the old transmales had a bone age of ≥14 years. The young transfemales group had a bone age of <15 years and the old transfemales group ≥15 years.		Change from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of ≥15; median [range]), GnRH analogue: 0.22 (0.18 to 0.25) g/cm3, gender-affirming hormones: 0.22 (0.19 to 0.24) g/cm3 (NS); z-score GnRH analogue: −1.18 (−1.78 to 1.09), gender-affirming hormones: −1.15 (−2.21 to 0.08) (p≤0.1) Change from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of <15 years; median [range]), GnRH analogue: 0.23 (0.20 to 0.29) g/cm3, gender-affirming hormones: 0.23 (0.19 to 0.28) g/cm3 (NS); z-score GnRH analogue: −0.05 (−0.78 to 2.94), gender-affirming hormones: −0.84 (−2.20 to 0.87) (p=0.003) Change from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of ≥15; median [range]), GnRH analogue: 0.26 (0.21 to 0.29) g/cm3, gender-affirming hormones: 0.24 (0.20 to 0.28) g/cm3 (p≤0.01); z-score GnRH analogue: 0.27 (−1.60 to 1.80), gender-affirming hormones: −0.29 (−2.28 to 0.90) (p≤ 0.0001) ***Bone density; femoral*** **Femoral neck BMAD** Change from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of <15 years; median [range]), GnRH analogue: 0.29 (0.20 to 0.33) g/cm3, gender-affirming hormones: 0.27 (0.20 to 0.33) g/cm3 (p≤0.1); z-score GnRH analogue: −0.71 (−3.35 to	3. Follow-up rate variable across outcomes and no description of those lost **Overall quality is assessed as poor**. Other comments: Within person comparison. No concomitant treatments were reported. Source of funding: grant from Abbott diagnostics

Study details	Population	Interventions	Study outcomes	Appraisal and Funding
			0.37), gender-affirming hormones: −1.32 (−3.39 to 0.21) (p≤0.1) Change from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of ≥15; median [range]), GnRH analogue: 0.30 (0.26 to 0.36) g/cm3, gender-affirming hormones: 0.30 (0.26 to 0.34) g/cm3 (NS); z-score GnRH analogue: −0.44 (−1.37 to 0.93), gender-affirming hormones: −0.36 (−1.50 to 0.46) (NS) Change from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of <15 years; median [range]), GnRH analogue: 0.31 (0.26 to 0.36) g/cm3, gender-affirming hormones: 0.30 (0.22 to 0.35) g/cm3 (NS); z-score GnRH analogue: −0.01 (−1.30 to 0.91), gender-affirming hormones: −0.37 (−2.28 to 0.47) (NS) Change from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of ≥15; median [range]), GnRH analogue: 0.33 (0.25 to 0.39) g/cm3, gender-affirming hormones: 0.30 (0.23 to 0.41) g/cm3 (p≤0.01); z-score GnRH analogue: 0.27 (−1.39 to 1.32), gender-affirming hormones: −0.27 (−1.91 to 1.29) (p=0.002)	

Appendix F Quality appraisal checklists

Newcastle-Ottawa tool for cohort studies

Question	
Domain: Selection	
1. Representativeness of the exposed cohort	Truly representative of the average [describe] in the community Somewhat representative of the average [describe] in the community Selected group of users e.g. nurses, volunteers No description of the derivation of the cohort
2. Selection of the non-exposed cohort	Drawn from the same community as the exposed cohort Drawn from a different source No description of the derivation of the non-exposed cohort
3. Ascertainment of exposure	Secure record (e.g. surgical records) Structured interview Written self-report No description
4. Demonstration that outcome of interest was not present at start of study	Yes / No
Domain: Comparability	
1. Comparability of cohorts on the basis of the design or analysis	Study controls for [select most important factor] Study controls for any additional factor [this criteria could be modified to indicate specific control for a second important factor]
Domain: Outcome	
1. Assessment of outcome	Independent blind assessment Record linkage Self-report No description
2. Was follow-up long enough for outcomes to occur	Yes [select and adequate follow up period for outcome of interest] No
3. Adequacy of follow up of cohorts	Complete follow up (all subjects accounted for) Subjects lost to follow up unlikely to introduce bias (small number lost to follow up [select an adequate %] follow up or description provided of those lost) Follow up rate [select an adequate %] and no description of those lost No statement

Appendix G Grade profiles

Table 2: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – gender dysphoria

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on gender dysphoria									
Mean±SD Utrecht Gender Dysphoria Scale[1] (version(s) not reported), time point at baseline (before GnRH analogues) versus follow-up (before gender-affirming hormones, higher scores indicate more gender dysphoria)									
1 cohort study de Vries et al 2011	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=41	None	Baseline: 53.20±7.91 GnRH analogue: 53.9±17.42 *P*=0.333	Critical	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; *P*, P-value; SD, Standard deviation.

1 The UGDS is a validated screening tool for both adolescents and adults to assess gender dysphoria. It consists of 12 items, to be answered on a 1- to 5-point scale, resulting in a sum score between 12 and 60. The higher the UGDS score the greater the gender dysphoria.
2 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).

Table 3: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – mental health

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on mental health									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Mean±SD Beck Depression Inventory-II, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones). (Lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=41	None	Baseline: 8.31±7.12 GnRH analogue: 4.95±6.72 *P*=0.004	Critical	VERY LOW
Mean±SD Trait Anger (TPI), time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=41	None	Baseline: 18.29±5.54 GnRH analogue: 17.88±5.24 *P*=0.503	Critical	VERY LOW
Mean±SD Trait Anxiety (STAI), time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=41	None	Baseline: 39.43±10.07 GnRH analogue: 37.95±9.38 *P*=0.276	Critical	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).

Table 4: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – body image

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Impact on body image									
Mean±SD Body Image Scale (primary sexual characteristics), time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=57	None	Baseline: 4.10±0.56 GnRH analogue: 3.98±0.71 *P*=0.145	Important	VERY LOW
Mean±SD Body Image Scale (secondary sexual characteristics), time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=57	None	Baseline: 2.74±0.65 GnRH analogue: 2.82±0.68 *P*=0.569	Important	VERY LOW
Mean±SD Body Image Scale (neutral characteristics), time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit)									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=57	None	Baseline: 2.41±0.63 GnRH analogue: 2.47±0.56 *P*=0.620	Important	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).

Table 5: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – psychosocial impact

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Psychosocial impact									
Mean [±SD] Children's Global Assessment Scale score, at baseline, higher scores indicate benefit)									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	n=101 58.72 [±11.38]	n=100 56.63 [±13.14]	*P*=0.23	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, at 6 months[2] (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	n=101 60.89 [±12.17]	n=100 60.29 [±12.81]	*P*=0.73	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, at 12 months[3] (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	n=60 64.70 [±13.34]	n=61 62.97 [±14.10]	*P*=0.49	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, at 18 months[4] (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	n=35 67.40 [±13.93]	n=36 62.53 [±13.54]	*P*=0.14	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 6 months compared to baseline (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=101 N=101	None	Baseline: 58.72±11.38 6 months: 60.89±12.17 *P*=0.19	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 12 months compared to baseline (higher scores indicate benefit).									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=101 N=60	None	Baseline: 58.72±11.38 12 months: 64.70±13.34 *P*=0.003	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 18 months compared to baseline (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=101 N=35	None	Baseline: 58.72±11.38 18 months: 67.40±13.93 *P*<0.001	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 12 months compared to 6 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=101 N=60	None	6 months: 60.89±12.17 12 months: 64.70±13.34 *P*=0.07	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 18 months compared to 6 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=101 N=35	None	6 months: 60.89±12.17 18 months: 67.40±13.93 *P*<0.001	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, participants at 18 months compared to 12 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=60 N=35	None	12 months: 64.70±13.34 18 months: 67.40±13.93 *P*=0.35	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 6 months[2] compared to baseline (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=201	None	Baseline: 57.73±12.27 6 months: 60.68±12.47 *P*<0.001	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 12 months[3] compared to baseline (higher scores indicate benefit).									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=201 N=121	None	Baseline: 57.73±12.27 12 months: 63.31±14.41 *P*<0.001	Important	VERY LOW
Mean±SD Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 18 months[4] compared to baseline (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=201 N=71	None	Baseline: 57.73±12.27 18 months: 64.93±13.85 *P*<0.001	Important	VERY LOW
Mean±SD Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 12 months compared to 6 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=201 N=121	None	6 months: 60.68±12.47 12 months: 63.31±14.41 *P*<0.08	Important	VERY LOW
Mean±SD Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 18 months compared to 6 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=201 N=71	None	6 months: 60.68±12.47 18 months: 64.93±13.85 *P*<0.02	Important	VERY LOW
Mean±SD Children's Global Assessment Scale score, in all participants (including those not treated with GnRH analogues) at 18 months compared to 12 months (higher scores indicate benefit).									
1 cohort study Costa et al 2015	Serious limitations[1]	No serious indirectness	No serious inconsistency	Not calculable	N=121 N=71	None	12 months: 63.31±14.41 18 months: 64.93±13.85 *P*<0.45	Important	VERY LOW
Mean±SD Children's Global Assessment Scale score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, higher scores indicate benefit).									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=41	None	Baseline: 70.24±10.12 GnRH analogue: 73.90±9.63 *P*=0.005	Important	VERY LOW
Mean±SD Child Behaviour Checklist (total T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 60.70±12.76 GnRH analogue: 54.46±11.23 *P*<0.001	Important	VERY LOW
Mean±SD Child Behaviour Checklist (internalising T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 61.00±12.21 GnRH analogue: 52.1±9.81 *P*<0.001	Important	VERY LOW
Mean±SD Child Behaviour Checklist (externalising T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 58.04±12.99 GnRH analogue: 53.81±11.86 *P*=0.001	Important	VERY LOW
Proportion of adolescents scoring in the clinical range Child Behaviour Checklist total problem scale, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 44.4% GnRH analogue: 22,2% *P*=0.001	Important	VERY LOW
Mean±SD Youth Self-Report (total T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormone, lower scores indicate benefit).									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 55.46±11.56 GnRH analogue: 50.00±10.56 *P*<0.001	Important	VERY LOW
Mean±SD Youth Self-Report (internalising T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 56.04±12.49 GnRH analogue: 49.78±11.63 *P*<0.001	Important	VERY LOW
Mean±SD Youth Self-Report (externalising T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 53.30±11.87 GnRH analogue: 49.98±9.35 *P*=0.009	Important	VERY LOW
Proportion of adolescents scoring in the clinical range Youth Self-Report (internalising T) score, time point at baseline (before GnRH analogues) versus follow-up (just before gender-affirming hormones, lower scores indicate benefit).									
1 cohort study de Vries et al 2011	Serious limitations[5]	No serious indirectness	Not applicable	Not calculable	N=54	None	Baseline: 29.6% GnRH analogue: 11.1% *P*=0.017	Important	VERY LOW
Mean±SD Child Behaviour Checklist score, transfemales (lower scores indicate benefit									
1 cross-sectional study Staphorsius et al 2015	Serious limitations[6]	No serious indirectness	Not applicable	Not calculable	N=8	N=10	GnRH analogue: 57.4 [±9.8] No GnRH analogue: 58.2 [±9.3]	Important	VERY LOW
Mean±SD Child Behaviour Checklist score, transmales (lower scores indicate benefit)									
1 cross-sectional study	Serious limitations[6]	No serious indirectness	Not applicable	Not calculable	N=12	N=10	GnRH analogues: 57.5 [±9.4] No GnRH analogue: 63.9 [±10.5]	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Staphorsius et al 2015									

Abbreviations: GnRH, gonadotrophin releasing hormone; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by Costa et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
2 6 months from baseline (after 6 months of psychological support – both groups).
3 12 months from baseline (delayed eligible gender dysphoria [GD] adolescents, after 12 months of psychological support; immediately eligible GD adolescents, after 12 months of psychological support + 6 months of puberty suppression).
4 18 months from baseline (delayed eligible gender dysphoria [GD] adolescents, after 12 months of psychological support; immediately eligible GD adolescents, after 12 months of psychological support + 6 months of puberty suppression).
5 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
6 Downgraded 1 level - the cohort study by Staphorsius et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding and no randomisation).

Table 6: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – engagement with healthcare services

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Engagement with healthcare services									
Number (proportion) failing to engage with health care services (did not attend clinic), at (up to) 9 years follow-up									
1 cohort study Brik et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	9/214 (4.2%)	None	9 adolescents out of 214 failed to attend clinic and were excluded from the study (4.2%)	Important	VERY LOW
Loss to follow-up									
1 cohort study	Serious limitations[2]	No serious indirectness	Not applicable		201	None	The sample size at baseline and 6 months was 201, which dropped by 39.8% to 121 after	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Costa et al 2015				Not calculable			12 months and by 64.7% to 71 at 18 months follow-up. No explanation of the reasons for loss to follow-up are reported.		

Abbreviations: GnRH, gonadotrophin releasing hormone.

1 Downgraded 1 level - the cohort study by Brik et al. (2018) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
2 Downgraded 1 level - the cohort study by Costa et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).

Table 7: Question 1. For children and adolescents with gender dysphoria, what is the clinical effectiveness of treatment with GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – stopping treatment

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Stopping treatment									
Number (proportion) stopping GnRH analogues, at (up to) 9 years follow-up									
1 cohort study Brik et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	9/143 (6.2%)	None	9/143 adolescents stopped GnRH analogues (6.2%)[2]	Important	VERY LOW
Number (proportion) stopping from GnRH analogues, at (up to) 13 years follow-up									
1 cohort study Khatchadourian et al 2014	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	11/27 (42%)	None	11/26 stopped GnRH analogues (42%)[4]	Important	VERY LOW
Number (proportion) stopping GnRH analogues but who wished to continue endocrine treatment, at (up to) 9 years follow-up									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 cohort study Brik et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	4/143 (2.8%)	None	4/143 adolescents stopped GnRH analogues but wished to continue treatment (2.8%)	Important	VERY LOW
Number (proportion) stopping GnRH analogues who no longer wished gender-affirming treatment, at (up to) 9 years follow-up									
1 cohort study Brik et al 2018	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	5/143 (3.5%)	None	5/143 adolescents stopped GnRH analogues and no longer wished to continue gender-affirming treatment (3.5%)	Important	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone.

1 Downgraded 1 level - the cohort study by Brik et al. (2018) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
2 Median duration of 0.8 years (range 0.1 to 3.0). Five adolescents stopped treatment because they no longer wished to receive gender-affirming treatment for various reasons. In 4 adolescents (all transmales), although they wanted to continue treatments for gender dysphoria, GnRH analogues were stopped mainly because of adverse effects (such as mood and emotional lability).
3 Downgraded 1 level - the cohort study by Khatchadourian et al. (2014) was assessed as at high risk of bias (poor quality overall; lack of blinding, no control group and high number of participants lost to follow-up).
4 Because of transitioning to gender-affirming hormones or gender-affirming surgery, adverse effects (such as mood and emotional lability) or no longer wishing to pursue transition.

Table 8. Question 2. For children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – bone density

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Bone density: change in lumbar BMAD									
Change in lumbar spine BMAD from baseline to 1 year in transfemales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=31	None	Mean (SD), g/cm^3 Baseline: 0.235 (0.030) 1 year: 0.233 (0.029) p=0.459 z-score Baseline: 0.859 (0.154) 1 year: −0.228 (1.027) p=0.000	IMPORTANT	VERY LOW
Change in lumbar spine BMAD from baseline to 1 year in transmales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=39	None	Mean (SD), g/cm^3 Baseline: 0.196 (0.035) 1 year: 0.201 (0.033) p=0.074 z-score Baseline: −0.186 (1.230) 1 year: −0.541 (1.396) p=0.006	IMPORTANT	VERY LOW
Change in lumbar spine BMAD from baseline to 2 years in transfemales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=10	None	Mean (SD), g/cm^3 Baseline: 0.240 (0.027) 2 years: 0.240 (0.030) p=0.865 z-score Baseline: 0.486 (0.809) 2 years: −0.279 (0.930) p=0.000	IMPORTANT	VERY LOW
Change in lumbar spine BMAD from baseline to 2 years in transmales									
1 observational study	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=21	None	Mean (SD), g/cm^3 Baseline: 0.195 (0.058) 2 years: 0.198 (0.055) p=0.433	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Joseph et al. (2019)							z-score Baseline: −0.361 (1.439) 2 years: −0.913 (1.318) p=0.001		
Change in lumbar BMAD from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=11 N=12	None	Mean (SD), g/cm^3 GnRH analogue: 0.22 (0.03) Gender-affirming hormones: 0.22 (0.02) NS z-score GnRH analogue: −0.44 (1.10) Gender-affirming hormones: −0.90 (0.80) p-value: NS	IMPORTANT	VERY LOW
Change in lumbar BMAD from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=18	None	Mean (SD), g/cm^3 GnRH analogue: 0.25 (0.03) Gender-affirming hormones: 0.24 (0.02) NS z-score GnRH analogue: 0.28 (0.90) Gender-affirming hormones: −0.50 (0.81) p-value: 0.004	IMPORTANT	VERY LOW
Change in lumbar BMAD from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of <15 years)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=15	None	Median (range), g/cm^3 GnRH analogue: 0.21 (0.17 to 0.25) Gender-affirming hormones: 0.20 (0.18 to 0.24)	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							NS z-score GnRH analogue: −0.20 (−1.82 to 1.18) Gender-affirming hormones: −1.52 (−2.36 to 0.42) p-value: <0.01		
Change in lumbar BMAD from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of ≥15)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=5	None	Median (range), g/cm^3 GnRH analogue: 0.22 (0.18 to 0.25) Gender-affirming hormones: 0.22 (0.19 to 0.24) NS z-score GnRH analogue: −1.18 (−1.78 to 1.09) Gender-affirming hormones: −1.15 (−2.21 to 0.08) p-value: p≤0.1	IMPORTANT	VERY LOW
Change in lumbar BMAD from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of <14 years)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=11	None	Median (range), g/cm^3 GnRH analogue: 0.23 (0.20 to 0.29) Gender-affirming hormones: 0.23 (0.19 to 0.28) NS z-score GnRH analogue: −0.05 (−0.78 to 2.94) Gender-affirming hormones: −0.84 (−2.20 to 0.87) p-value: ≤0.01	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Change in lumbar BMAD from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of ≥14)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=23	None	Median (range), g/cm3 GnRH analogue: 0.26 (0.21 to 0.29) Gender-affirming hormones: 0.24 (0.20 to 0.28) p≤0.01 z-score GnRH analogue: 0.27 (−1.60 to 1.80) Gender-affirming hormones: −0.29 (−2.28 to 0.90) p-value: p ≤ 0.01)	IMPORTANT	VERY LOW
Bone density: change in lumbar BMD									
Change in lumbar spine BMD from baseline to 1 year in transfemales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=31	None	Mean (SD), kg/m2 Baseline: 0.860 (0.154) 1 year: 0.859 (0.129) p=0.962 z-score Baseline: −0.016 (1.106) 1 year: −0.461 (1.121) p=0.003	IMPORTANT	VERY LOW
Change in lumbar spine BMD from baseline to 1 year in transmales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=39	None	Mean (SD), kg/m2 Baseline: 0.694 (0.149) 1 year: 0.718 (0.124) p=0.006 z-score Baseline: −0.395 (1.428) 1 year: −1.276 (1.410) p=0.000	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Change in lumbar spine BMD from baseline to 2 years in transfemales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=10	None	Mean (SD), kg/m2 Baseline: 0.867 (0.141) 2 years: 0.878 (0.130) p=0.395 z-score Baseline: 0.130 (0.972) 2 years: −0.890 (1.075) p=0.000	IMPORTANT	VERY LOW
Change in lumbar spine BMD from baseline to 2 years in transmales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=21	None	Mean (SD), kg/m2 Baseline: 0.695 (0.220) 2 years: 0.731 (0.209) p=0.058 z-score Baseline: −0.715 (1.406) 2 years: −2.000 (1.384) p=0.000	IMPORTANT	VERY LOW
Change in lumbar BMD from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=12 N=11	None	Mean (SD), g/m2 GnRH analogue: 0.84 (0.13) Gender-affirming hormones: 0.84 (0.11) NS z-score GnRH analogue: −0.77 (0.89) Gender-affirming hormones: −1.01 (0.98) NS	IMPORTANT	VERY LOW
Change in lumbar BMD from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=18	None	Mean (SD), g/m2 GnRH analogue: 0.95 (0.12) Gender-affirming hormones: 0.91 (0.10) p-value: 0.006 z-score GnRH analogue: 0.17 (1.18) Gender-affirming hormones: −0.72 (0.99) p-value: <0.001	IMPORTANT	VERY LOW
Bone density: change in femoral neck (hip) BMD									
Change in femoral neck BMD from baseline to 1 year in transfemales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=31	None	Mean (SD), kg/m2 Baseline: 0.894 (0.118) 1 year: 0.905 (0.104) p=0.571 z-score Baseline: 0.157 (0.905) 1 year: −0.340 (0.816) p=0.002	IMPORTANT	VERY LOW
Change from baseline to 1 year in femoral neck BMD in transmales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=39	None	Mean (SD), kg/m2 Baseline: 0.772 (0.137) 1 year: 0.785 (0.120) p=0.797 z-score Baseline: −0.863 (1.215) 1 year: −1.440 (1.075) p=0.000	IMPORTANT	VERY LOW
Change from baseline to 2 years in femoral neck BMD in transfemales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=10	None	Mean (SD), kg/m2 Baseline: 0.920 (0.116) 2 years: 0.910 (0.125) p=0.402 z-score Baseline: 0.450 (0.781) 2 years: −0.600 (1.059) p=0.002	IMPORTANT	VERY LOW
Change from baseline to 2 years in femoral neck BMD in transmales									
1 observational study Joseph et al. (2019)	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=21	None	Mean (SD), kg/m2 Baseline: 0.766 (0.215) 2 years: 0.773 (0.197) p=0.604 z-score Baseline: −1.075 (1.145) 2 years: −1.779 (0.816) p=0.001	IMPORTANT	VERY LOW
Bone density: change in femoral neck (hip) BMAD									
Change from starting GnRH analogue to starting gender-affirming hormones in femoral neck BMAD in transfemales (bone age of <15 years)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=16	None	Median (range), g/cm3 GnRH analogue: 0.29 (0.20 to 0.33) Gender-affirming hormones: 0.27 (0.20 to 0.33) p≤0.1 z-score GnRH analogue: −0.71 (−3.35 to 0.37) Gender-affirming hormones: −1.32 (−3.39 to 0.21) p≤0.1	IMPORTANT	VERY LOW
Change in femoral neck BMAD from starting GnRH analogue to starting gender-affirming hormones in transfemales (bone age of ≥15)									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=6	None	Median (range), g/cm3 GnRH analogue: 0.30 (0.26 to 0.36) Gender-affirming hormones: 0.30 (0.26 to 0.34) NS z-score GnRH analogue: −0.44 (−1.37 to 0.93) Gender-affirming hormones: −0.36 (−1.50 to 0.46) NS	IMPORTANT	VERY LOW
Change in femoral neck BMAD from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of <14 years)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=10	None	Median (range), g/cm3 GnRH analogue: 0.31 (0.26 to 0.36) Gender-affirming hormones: 0.30 (0.22 to 0.35) NS z-score GnRH analogue: −0.01 (−1.30 to 0.91) Gender-affirming hormones: −0.37 (−2.28 to 0.47) NS	IMPORTANT	VERY LOW
Change in femoral neck BMAD from starting GnRH analogue to starting gender-affirming hormones in transmales (bone age of ≥14)									
1 observational study Vlot et al. 2017	Serious limitations[3]	No serious indirectness	Not applicable	Not calculable	N=23	None	Median (range), g/cm3 GnRH analogue: 0.33 (0.25 to 0.39) Gender-affirming hormones: 0.30 (0.23 to 0.41) p-value: ≤0.01 z-score	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
							GnRH analogue: 0.27 (−1.39 to 1.32) Gender-affirming hormones: −0.27 (−1.91 to 1.29) p-value: ≤0.01		
Bone density: change in femoral area BMD									
Change in femoral BMD from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=14 N=6	None	Mean (SD), g/m2 GnRH analogue: 0.88 (0.12) Gender-affirming hormones: 0.87 (0.08) NS z-score GnRH analogue: −0.66 (0.77) Gender-affirming hormones: −0.95 (0.63) NS	IMPORTANT	VERY LOW
Change in femoral BMD from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=18 N=13	None	Mean (SD), g/m2 GnRH analogue: 0.92 (0.10) Gender-affirming hormones: 0.88 (0.09) p-value: 0.005 z-score GnRH analogue: 0.36 (0.88) Gender-affirming hormones: −0.35 (0.79) p-value: 0.001	IMPORTANT	VERY LOW
Bone density: change in femoral area BMAD									
Change in femoral BMAD from starting GnRH analogue (mean age 14.9±1.9) to starting gender-affirming hormones (mean age 16.6±1.4) in transfemales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=12 N=10	None	Mean (SD), g/cm3 GnRH analogue: 0.28 (0.04) Gender-affirming hormones: 0.26 (0.04) NS z-score GnRH analogue: −0.93 (1.22) Gender-affirming hormones: −1.57 (1.74) p-value: NS	IMPORTANT	VERY LOW
Change in femoral BMAD from starting GnRH analogue (mean age 15.0±2.0) to starting gender-affirming hormones (mean age 16.4±2.3) in transmales									
1 observational study Klink et al. 2015	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable	N=18 N=18	None	Mean (SD), g/cm3 GnRH analogue: 0.32 (0.04) Gender-affirming hormones: 0.31 (0.04) NS z-score GnRH analogue: 0.01 (0.70) Gender-affirming hormones: −0.28 (0.74) NS	IMPORTANT	VERY LOW

Abbreviations: BMAD, bone mineral apparent density; BMD, bone mineral density; GnRH, gonadotrophin releasing hormone; NS, not significant; SD, standard deviation.

1 Downgraded 1 level - the cohort study by Joseph et al. (2019) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).

2 Downgraded 1 level - the cohort study by Klink et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding, no randomisation, no control group and high number of participants lost to follow-up).

3 Downgraded 1 level - the cohort study by Vlot et al. (2017) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control).

Table 9 Question 2: For children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – cognitive development or functioning

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Cognitive development or functioning (1 cross-sectional study)									
IQ (4 subscales: arithmetic, vocabulary, picture arrangement, and block design) at a single time point between GnRH analogue treated and untreated transfemales									
1 Cross-sectional study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=8 Mean (SD) 94.0 (10.3)	N=10 Mean (SD) 109.4 (21.2)	NR	IMPORTANT	VERY LOW
IQ (4 subscales: arithmetic, vocabulary, picture arrangement, and block design) at a single time point between GnRH analogue treated and untreated transmales									
1 Cross-sectional study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=12 Mean (SD) 95.8 (15.6)	N=10 Mean (SD) 98.5 (15.9)	NR	IMPORTANT	VERY LOW
Reaction time at a single time point between GnRH analogue treated and untreated transfemales									
1 Cross-sectional study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=8 Mean (SD) 10.9 (4.1)	N=10 Mean (SD) 9.9 (3.1)	NR	IMPORTANT	VERY LOW
Reaction time at a single time point between GnRH analogue treated and untreated transmales									
1 Cross-sectional study	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=12 Mean (SD) 9.9 (3.1)	N=10 Mean (SD) 10.0 (2.0)	NR	IMPORTANT	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Staphorsius et al. 2015									
Accuracy at a single time point between GnRH analogue treated and untreated transfemales									
1 cohort study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=8 Mean (SD) 73.9 (9.1)	N=10 Mean (SD) 83.4 (9.5)	NR	IMPORTANT	VERY LOW
Accuracy at a single time point between GnRH analogue treated and untreated transmales									
1 cohort study Staphorsius et al. 2015	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=12 Mean (SD) 85.7 (10.5)	N=10 Mean (SD) 88.8 (9.7)	NR	IMPORTANT	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; NR, not reported; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by Staphorsius et al. (2015) was assessed as at high risk of bias (poor quality overall; lack of blinding and no randomisation).

Table 10: Question 2: In children and adolescents with gender dysphoria, what is the short-term and long-term safety of GnRH analogues compared with one or a combination of psychological support, social transitioning to the desired gender or no intervention? – other safety outcomes

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
Other safety outcomes: change in serum creatinine									
Change in serum creatinine (micromol/l) between baseline and 1 year in transfemales									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients% (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Intervention	Comparator	Result		
1 observational study Schagen et al. 2016	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=28	None	Mean (SD) Baseline: 70 (12) 1 year: 66 (13) p-value: 0.20	IMPORTANT	VERY LOW
Change in serum creatinine (µmol/l) between baseline and 1 year in transmales									
1 observational study Schagen et al. 2016	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	N=29	None	Mean (SD) Baseline: 73 (8) 1 year: 68 (13) p-value: 0.01	IMPORTANT	VERY LOW
Other safety outcomes: liver enzymes									
Presence of elevated liver enzymes (AST, ALT, and glutamyl transferase) between baseline and during treatment									
1 observational study Schagen et al. 2016	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	39	None	Glutamyl transferase was not elevated at baseline or during treatment in any subject. Mild elevations of AST and ALT above the reference range were present at baseline but were not more prevalent during treatment than at baseline. Glutamyl transferase, AST, and ALT levels did not significantly change from baseline to 12 months of treatment.	IMPORTANT	VERY LOW
Other safety outcomes: adverse effects									
Proportion of patients reporting adverse effects									
1 cohort study Khatchadourian et al 2014	Serious limitations[2]	No serious indirectness	Not applicable	Not calculable[2]	27	None	3/27 adolescents[3]	Important	VERY LOW

Abbreviations: ALT, alanine aminotransferase; AST, aspartate aminotransferase; GnRH, gonadotrophin releasing hormone; P, P-value; SD, standard deviation.

1 Downgraded 1 level - the cohort study by Schagen et al. (2016) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control).

2 Downgraded 1 level - the cohort study by Khatchadourian et al. (2014) was assessed as at high risk of bias (poor quality overall; lack of blinding, no control group and high number of participants lost to follow-up).

3 1 transmale developed sterile abscesses; they were switched from leuprolide acetate to triptorelin, and this was well tolerated. 1 transmale developed leg pains and headaches, which eventually resolved without treatment. 1 participant gained 19 kg within 9 months of initiating GnRH analogues.

Table 11: Question 4. From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may derive more (or less) advantage from treatment with GnRH analogues than the wider population of children and adolescents with gender dysphoria? – critical outcomes

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
Subgroups: sex assigned at birth males compared with sex assigned at birth females									
Impact on gender dysphoria									
Mean [±SD] Utrecht Gender Dysphoria Scale (version(s) not reported), time point at baseline (before GnRHa) versus follow-up (just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 47.95 [±9.70] score at T1 49.67 [±9.47]	n-NR[2] score at T0 56.57 [±3.89] score at T1 56.62 [±4.0]	*F*-ratio 15.98 (*df, errdf*: 1,39), *P*<0.001	Critical	VERY LOW
Impact on mental health									
Mean [±SD] Beck Depression Inventory-II, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 5.71 [±4.31] score at T1 3.50 [±4.58]	n-NR[2] score at T0 10.34 [±8.24] score at T1 6.09 [±7.93]	*F*-ratio 3.85 (*df, errdf*: 1,39), *P*=0.057	Critical	VERY LOW
Mean [±SD] Trait Anger (TPI), time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 5.22 [±2.76] score at T1 5.00 [±3.07]	n-NR[2] score at T0 6.43 [±2.78] score at T1 6.39 [±2.59]	*F*-ratio 5.70 (*df, errdf*: 1,39), *P*=0.022	Critical	VERY LOW
Mean [±SD] Trait Anxiety (STAI), time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 4.33 [±2.68] score at T1 4.39 [±2.64]	n-NR[2] score at T0 7.00 [±2.36] score at T1 6.17 [±2.69]	*F*-ratio 16.07 (*df, errdf*: 1,39), *P*<0.001	Critical	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; NR, not reported; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
2 The overall sample size completing the outcome at both time points was 41.

Table 11: Question: 4. From the evidence selected, are there any subgroups of children and adolescents with gender dysphoria that may derive more (or less) advantage from treatment with GnRH analogues than the wider population of children and adolescents with gender dysphoria? – important outcomes

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
Subgroups: sex assigned at birth males compared with sex assigned at birth females									
Impact on body image									
Mean [±SD] Body Image Scale (primary sexual characteristics), time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 4.02 [±0.16] score at T1 3.74 [±0.78]	n-NR[2] score at T0 4.16 [±0.52] score at T1 4.17 [±0.58]	*F*-ratio 4.11 (*df, errdf*: 1,55), *P*=0.047	Important	VERY LOW
Mean [±SD] Body Image Scale (secondary sexual characteristics), time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 2.66 [±0.50] score at T1 2.39 [±0.69]	n-NR[2] score at T0 2.81 [±0.76] score at T1 3.18 [±0.42]	*F*-ratio 11.57 (*df, errdf*: 1,55), *P*=0.001[3]	Important	VERY LOW
Mean [±SD] Body Image Scale (neutral characteristics), time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[2] score at T0 2.60 [±0.58] score at T1 2.32 [±0.59]	n-NR[2] score at T0 2.24 [±0.62] score at T1 2.61 [±0.50]	*F*-ratio 0.081 (*df, errdf*: 1,55), *P*=0.777[3]	Important	VERY LOW
Psychosocial impact									
Mean [±SD] Children's Global Assessment Scale score, at baseline.									
1 cohort study Costa et al 2015	Serious limitations[4]	No serious indirectness	No serious inconsistency	Not calculable	n=not reported 55.4 [±12.7]	n=not reported 59.2 [±11.8]	*t*-test 2.15; *P*=0.03[5]	Important	VERY LOW
Mean [±SD] Children's Global Assessment Scale score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[6] score at T0 73.10 [±8.84] score at T1 77.33 [±8.69]	n-NR[6] score at T0 67.25 [±11.06] score at T1 70.30 [±9.44]	*F*-ratio 5.77 (*df, errdf*: 1,39), *P*=0.021	Important	VERY LOW
Mean [±SD] Child Behaviour Checklist (total T) score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 59.42 [±11.78] score at T1 50.38	n-NR[7] score at T0 61.73 [±13.60]	*F*-ratio 2.64 (*df, errdf*: 1,52), *P=0.110*	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
					[±10.57]	score at T1 57.73 [±10.82]			
Mean [±SD] Child Behaviour Checklist (internalising T) score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 60.00 [±9.51] score at T1 52.17 [±9.81]	n-NR[7] score at T0 61.80 [±14.12] score at T1 56.30 [±10.33]	*F*-ratio 1.16 (*df, errdf*: 1,52), *P*=0.286	Important	VERY LOW
Mean [±SD] Child Behaviour Checklist (externalising T) score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 54.71 [±12.91] score at T1 48.75 [±10.22]	n-NR[7] score at T0 60.70 [±12.64] score at T1 57.87 [±11.66]	*F*-ratio 6.29 (*df, errdf*: 1,52), *P*=0.015	Important	VERY LOW
Mean [±SD] Youth Self-Report (total T) score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 53.56 [±12.26] score at T1 47.84 [±10.86]	n-NR[7] score at T0 57.10 [±10.87] score at T1 51.86 [±10.11]	*F*-ratio 1.99 (*df, errdf*: 1,52), *P*=0.164	Important	VERY LOW

QUALITY					Summary of findings			IMPORTANCE	CERTAINTY
					No of events/No of patients (n/N%)		Effect		
Study	Risk of bias	Indirectness	Inconsistency	Imprecision	Sex assigned at birth males	Sex assigned at birth females	Result		
Mean [±SD] Youth Self-Report (internalising T) score, time point at baseline (T0 before GnRH analogues) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 55.88 [±11.81] score at T1 49.24 [±12.24]	n-NR[7] score at T0 56.17 [±13.25] score at T1 50.24 [±11.28]	*F*-ratio 0.049 (*df, errdf*: 1,52), *P*=0.825	Important	VERY LOW
Mean [±SD] Youth Self-Report (externalising T) score, time point at baseline (T0 before GnRHa) versus follow-up (T1 just before gender-affirming hormones).									
1 cohort study de Vries et al 2011	Serious limitations[1]	No serious indirectness	Not applicable	Not calculable	n-NR[7] score at T0 48.72 [±11.83] score at T1 46.52 [±9.23]	n-NR[7] score at T0 57.24 [±10.59] score at T1 52.97 [±8.51]	*F*-ratio 9.14 (*df, errdf*: 1,52), *P*=0.004	Important	VERY LOW

Abbreviations: GnRH, gonadotrophin releasing hormone; NR, not reported; *P*, P-value; SD, Standard deviation.

1 Downgraded 1 level - the cohort study by de Vries et al. (2011) was assessed as at high risk of bias (poor quality overall; lack of blinding and no control group).
2 The overall sample size completing the outcome at both time points was 57.
3 There was a significant interaction effect between sex assigned at birth and BDI between T0 and T1; sex assigned at birth females became more dissatisfied with their secondary F (df, errdf), P: 14.59 (1,55), P<0.001) and neutral F (df, errdf), P: 15.26 (1,55), P<0.001) sex characteristics compared with sex assigned at birth males.
4 Serious limitations – the cohort study by Costa et al. 2015 was assessed as at high risk of bias (poor quality).
5 At baseline, CGAS scores were not associated with any demographic variable, in both sex assigned at birth males and females. There were no statistically significant differences in CGAS scores between gender dysphoric sex assigned at birth males and females in all follow-up evaluations (P>0.1; full data not reported).
6 The overall sample size completing the outcome at both time points was 41
7 The overall sample size completing the outcome at both time points was 54.

Glossary

Beck Depression Inventory-II (BDI-II)	The BDI-II is a tool for assessing depressive symptoms. There are no specific scores to categorise depression severity, but it is suggested that 0 to 13 is minimal symptoms, 14 to 19 is mild depression, 20 to 28 is moderate depression, and severe depression is 29 to 63.
Body Image Scale (BIS)	The BIS is used to measure body satisfaction. The scale consists of 30 body features, which the person rates on a 5-point scale. Each of the 30 items falls into one of 3 basic groups based on its relative importance as a gender-defining body feature: primary sex characteristics, secondary sex characteristics, and neutral body characteristics. A higher score indicates more dissatisfaction.
Bone mineral apparent density (BMAD)	BMAD is a size adjusted value of bone mineral density (BMD) incorporating body size measurements using UK norms in growing adolescents.
Child Behaviour Checklist (CBCL)	CBCL is a checklist parents complete to detect emotional and behavioural problems in children and adolescents.
Children's Global Assessment Scale (CGAS)	The CGAS tool is a validated measure of global functioning on a single rating scale from 1 to 100. Lower scores indicate poorer functioning.
Gender	The roles, behaviours, activities, attributes, and opportunities that any society considers appropriate for girls and boys, and women and men.
Gender dysphoria	Discomfort or distress that is caused by a discrepancy between a person's gender identity (how they see themselves regarding their gender) and that person's sex assigned at birth (and the associated gender role, and/or primary and secondary sex characteristics).
Gonadotrophin releasing hormone (GnRH) analogues	GnRH analogues competitively block GnRH receptors to prevent the spontaneous release of 2 gonadotropin hormones, Follicular Stimulating Hormone (FSH) and Luteinising Hormone (LH) from the pituitary gland. The reduction in FSH and LH secretion reduces oestradiol secretion from the ovaries in those whose sex assigned at birth was female and testosterone secretion from the testes in those whose sex assigned at birth was male.
Sex assigned at birth	Sex assigned at birth (male or female) is a biological term and is based on genes and how external and internal sex and reproductive organs work and respond to hormones. Sex is the label that is recorded when a baby's birth is registered.
Tanner stage	Tanner staging is a scale of physical development.
Trait Anger Spielberger scales of the State-Trait Personality Inventory (TPI)	The TPI is a validated 20-item inventory tool which measures the intensity of anger as the disposition to experience angry feelings as a personality trait. Higher scores indicate greater anger.
Transgender (including transmale and transfemale)	Transgender is a term for someone whose gender identity is not congruent with their birth-registered sex. A transmale is a person who identifies as male and a transfemale is a person who identifies as female.

Utrecht Gender Dysphoria Scale (UGDS)	The UGDS is a validated screening tool for both adolescents and adults to assess gender dysphoria. It consists of 12 items, to be answered on a 1- to 5-point scale, resulting in a sum score between 12 and 60. The higher the UGDS score the greater the impact on gender dysphoria.
Youth Self-Report (YSR)	The self-administered YSR is a checklist to detect emotional and behavioural problems in children and adolescents. It is self-completed by the child or adolescent. The scales consist of a Total problems score, which is the sum of the scores of all the problem items. An internalising problem scale sums the anxious/depressed, withdrawn-depressed, and somatic complaints scores while the externalising problem scale combines rule-breaking and aggressive behaviour.

References

Included studies

- Brik T, Vrouenraets L, de Vries M et al. (2020). Trajectories of Adolescents Treated with Gonadotropin-Releasing Hormone Analogues for Gender Dysphoria. Archives of Sexual Behaviour. [Accessed 6 August 2020]
- Costa R, Dunsford M, Skagerberg E et al. (2015) Psychological Support, Puberty Suppression, and Psychosocial Functioning in Adolescents with Gender Dysphoria. Journal of Sexual Medicine. [online] Volume 12(11), Pages 2206-2214. Available at: https://doi.org/10.1111/jsm.13034 [Accessed 7 August 2020]
- de Vries A, Steensma T, Doreleijers T et al. (2011) Puberty Suppression in Adolescents with Gender Identity Disorder: A Prospective Follow-Up Study. The Journal of Sexual Medicine Volume 8, Issue 8, August, Pages 2276-2283. [Accessed 11 August 2020].
- Joseph T, Ting J, Butler G (2019) The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort. Journal of pediatric endocrinology & metabolism 32(10): 1077-1081
- Khatchadourian K, Shazhan A, Metzger D. (2014) Clinical Management of Youth with Gender Dysphoria in Vancouver. The Journal of Pediatrics. Volume 164, Issue 4, April, Pages 906-911. [Accessed 14 August 2020]
- Klink D, Caris M, Heijboer A et al. (2015) Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. The Journal of clinical endocrinology and metabolism 100(2): e270-5
- Schagen SEE, Cohen-Kettenis PT, Delemarre-van de Waal HA et al. (2016) Efficacy and Safety of Gonadotropin-Releasing Hormone Agonist Treatment to Suppress Puberty in Gender Dysphoric Adolescents. The journal of sexual medicine 13(7): 1125-32
- Staphorsius A, Baudewijntje P, Kreukels P, et al. (2015) Puberty suppression and executive functioning: An fMRI-study in adolescents with gender dysphoria. Psychoneuroendocrinology Volume 565. Pages 190-199. [Accessed 10 August 2020]

- Vlot, Mariska C, Klink, Daniel T, den Heijer, Martin et al. (2017) Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone 95: 11-19

Other references

- World Health Organisation (2018) International Classification of Diseases 11. Available from https://icd.who.int/ [online; accessed 20 August 2020]
- American Psychiatric Association. (2013). Diagnostic and statistical Manual of Mental Disorders (DSM-5) (5th ed). Washington, DC and London: American Psychiatric Publishing. pp.451-460. [accessed 20 August 2020]
- NHS England (2013). NHS Standard contract for gender identity development service for children and adolescents [accessed 20 August 2020]

Copyright

abbvie

PRESS RELEASE

AbbVie Reports Full-Year and Fourth-Quarter 2021 Financial Results

- *Reports Full-Year Diluted EPS of $6.45 on a GAAP Basis, an Increase of 137.1 Percent; Adjusted Diluted EPS of $12.70, an Increase of 20.3 Percent*
- *Delivers Full-Year Net Revenues of $56.197 Billion on a GAAP Basis, an Increase of 22.7 Percent; Adjusted Net Revenues Were $56.122 Billion*
- *Full-Year Global Net Revenues from the Immunology Portfolio Were $25.284 Billion, an Increase of 14.1 Percent on a Reported Basis, or 13.5 Percent on an Operational Basis; U.S. Humira Net Revenues Were $17.330 Billion, an Increase of 7.6 Percent; Internationally, Humira Net Revenues Were $3.364 Billion, a Decrease of 9.6 Percent on a Reported Basis, or 12.8 Percent on an Operational Basis, Due to Biosimilar Competition; Global Skyrizi Net Revenues Were $2.939 Billion; Global Rinvoq Net Revenues Were $1.651 Billion*
- *Full-Year Global Net Revenues from the Hematologic Oncology Portfolio Were $7.228 Billion, an Increase of 8.7 Percent on a Reported Basis, or 8.3 Percent on an Operational Basis; Global Imbruvica Net Revenues Were $5.408 Billion, an Increase of 1.8 Percent, with U.S. Net Revenues of $4.321 Billion and International Profit Sharing of $1.087 Billion; Global Venclexta Net Revenues Were $1.820 Billion*
- *Full-Year Global Net Revenues from the Neuroscience Portfolio Were $5.927 Billion; Global Botox Therapeutic Net Revenues Were $2.451 Billion; Vraylar Net Revenues Were $1.728 Billion*
- *Full-Year Global Net Revenues from the Aesthetics Portfolio Were $5.233 Billion; Global Botox Cosmetic Net Revenues Were $2.232 Billion*
- *Reports Fourth-Quarter Diluted EPS of $2.26 on a GAAP Basis, an Increase of Over 100.0 Percent; Adjusted Diluted EPS of $3.31, an Increase of 13.4 Percent*
- *Delivers Fourth-Quarter Net Revenues of $14.886 Billion, an Increase of 7.4 Percent on a GAAP Basis*
- *Provides 2022 GAAP Diluted EPS Guidance Range of $9.26 to $9.46; Provides 2022 Adjusted Diluted EPS Guidance Range of $14.00 to $14.20*

NORTH CHICAGO, Ill., February 2, 2022 – AbbVie (NYSE:ABBV) announced financial results for the fourth quarter and full year ended December 31, 2021.

"We delivered another year of outstanding performance in 2021 with double-digit revenue and EPS growth that were well above our initial expectations," said Richard A. Gonzalez, chairman and chief executive officer, AbbVie. "We are entering 2022 with significant momentum and expect our diverse set of growth assets, robust pipeline and excellent execution to deliver continued strong performance this year and over the long term."

Note: "Operational" comparisons are presented at constant currency rates that reflect comparative local currency net revenues at the prior year's foreign exchange rates.

Fourth-Quarter Results

- Worldwide net revenues were $14.886 billion, an increase of 7.4 percent on a reported basis, or 7.5 percent on an operational basis.
- Global net revenues from the immunology portfolio were $6.746 billion, an increase of 13.2 percent on a reported basis, or 13.3 percent on an operational basis.
  - Global Humira net revenues of $5.334 billion increased 3.5 percent on a reported and operational basis. U.S. Humira net revenues were $4.553 billion, an increase of 6.0 percent. Internationally, Humira net revenues were $781 million, a decrease of 9.1 percent on a reported basis, or 8.8 percent on an operational basis, due to biosimilar competition.
  - Global Skyrizi net revenues were $895 million.
  - Global Rinvoq net revenues were $517 million.
- Global net revenues from the hematologic oncology portfolio were $1.873 billion, an increase of 4.6 percent on a reported basis, or 4.7 percent on an operational basis.
  - Global Imbruvica net revenues were $1.385 billion, a decrease of 2.7 percent, with U.S. net revenues of $1.114 billion and international profit sharing of $271 million.
  - Global Venclexta net revenues were $488 million, an increase of 33.3 percent on a reported basis, or 34.0 percent on an operational basis.
- Global net revenues from the neuroscience portfolio were $1.654 billion, an increase of 19.0 percent on a reported and operational basis.
  - Global Botox Therapeutic net revenues were $671 million, an increase of 18.3 percent on a reported basis, or 18.1 percent on an operational basis.
  - Vraylar net revenues were $489 million, an increase of 21.8 percent.
  - Global Ubrelvy net revenues were $183 million.
- Global net revenues from the aesthetics portfolio were $1.407 billion, an increase of 23.3 percent on a reported basis, or 22.8 percent on an operational basis.
  - Global Botox Cosmetic net revenues were $626 million, an increase of 27.0 percent on a reported basis, or 26.6 percent on an operational basis.
  - Global Juvederm net revenues were $432 million, an increase of 30.6 percent on a reported basis, or 29.8 percent on an operational basis.
- On a GAAP basis, the gross margin ratio in the fourth quarter was 71.0 percent. The adjusted gross margin ratio was 83.6 percent.
- On a GAAP basis, selling, general and administrative expense was 21.9 percent of net revenues. The adjusted SG&A expense was 22.2 percent of net revenues.
- On a GAAP basis, research and development expense was 12.3 percent of net revenues. The adjusted R&D expense was 12.1 percent of net revenues, reflecting funding actions supporting all stages of our pipeline.
- On a GAAP basis, the operating margin in the fourth quarter was 34.1 percent. The adjusted operating margin was 49.3 percent.
- On a GAAP basis, net interest expense was $571 million.
- On a GAAP basis, the tax rate in the quarter was 5.3 percent. The adjusted tax rate was 12.5 percent.
- Diluted EPS in the fourth quarter was $2.26 on a GAAP basis. Adjusted diluted EPS, excluding specified items, was $3.31.

Note: "Operational" comparisons are presented at constant currency rates that reflect comparative local currency net revenues at the prior year's foreign exchange rates.

Recent Events

- AbbVie confirmed prior revenue guidance of greater than $15 billion in combined Skyrizi (risankizumab) and Rinvoq (upadacitinib) risk-adjusted sales in 2025. AbbVie expects each asset to deliver risk-adjusted sales of greater than $7.5 billion in 2025. Skyrizi is part of a collaboration between Boehringer Ingelheim and AbbVie, with AbbVie leading development and commercialization globally.

- AbbVie announced that the U.S. Food and Drug Administration (FDA) approved Rinvoq for the treatment of moderate to severe atopic dermatitis (AD) in adults and children 12 years of age and older whose disease did not respond to previous treatment and is not well controlled with other pills or injections, including biologic medicines, or when use of other pills or injections is not recommended. The approval includes two dose strengths (15 mg and 30 mg, once daily) and is supported by efficacy and safety data from one of the largest registrational Phase 3 programs for AD with more than 2,500 patients evaluated across three studies. This milestone marked the third FDA-approved indication for Rinvoq.

- AbbVie announced the FDA approved Rinvoq (15 mg, once daily) for the treatment of adults with active psoriatic arthritis (PsA) who have had an inadequate response or intolerance to one or more tumor necrosis factor (TNF) blockers. The approval is supported by two Phase 3 clinical studies where Rinvoq showed efficacy across multiple measures of disease activity in active PsA with a safety profile consistent with that seen in rheumatoid arthritis (RA). This milestone marked the second FDA-approved indication for Rinvoq.

- AbbVie announced the FDA approved Skyrizi for the treatment of adults with active PsA. The approval is supported by two Phase 3 clinical studies where Skyrizi demonstrated significant improvement in joint symptoms, including swollen, tender and painful joints, compared to placebo. This milestone marked the second FDA-approved indication for Skyrizi.

- AbbVie announced the European Commission (EC) approved Skyrizi alone or in combination with methotrexate (MTX), for the treatment of active PsA in adults who have had an inadequate response or who have been intolerant to one or more disease-modifying anti-rheumatic drugs (DMARDs). The positive opinion is based on data from two pivotal Phase 3 studies which evaluated the efficacy and safety of Skyrizi in adults with active PsA and marks Skyrizi's second indication in the European Union (EU).

- AbbVie announced that it submitted applications to the FDA and European Medicines Agency (EMA) seeking approval for Rinvoq (15 mg, once daily) for the treatment of adults with active non-radiographic axial spondyloarthritis (nr-axSpA). The submissions are supported by the Phase 3 SELECT-AXIS 2 (study 2) clinical trial in which Rinvoq demonstrated significant improvements in signs and symptoms as well as physical function and disease activity versus placebo. No new safety risks were observed compared to the known safety profile of Rinvoq. In addition, AbbVie requested label enhancements for Rinvoq in the EU to include adult patients with active AS who had an inadequate response to biologic DMARDs, based on newly generated clinical data. These data were also provided to the FDA in support of the agency's ongoing review of the supplemental New Drug Application (sNDA) for Rinvoq in AS.

- AbbVie announced that it submitted an application to the EMA seeking approval for Skyrizi (600 mg intravenous induction and 360 mg subcutaneous maintenance therapy) for the treatment of patients 16 years and older with moderate to severe Crohn's disease (CD). The submission is supported by three pivotal Phase 3 studies in which Skyrizi demonstrated significant improvements in clinical remission and endoscopic response as both induction and maintenance therapy. The overall safety findings in these pivotal studies were generally consistent with the known safety profile of Skyrizi. If approved, CD will mark the third indication for Skyrizi in the EU.

- AbbVie announced positive top-line results from the Phase 3 induction study, U-EXCEED, which showed Rinvoq (45 mg, once daily) achieved both primary endpoints of clinical remission and endoscopic response at week 12 as well as key secondary endpoints in patients with moderate to severe CD. The safety results in this study were consistent with the known profile of Rinvoq, with no new safety risks observed. U-EXCEED is the first of two Phase 3 induction studies to evaluate the safety and efficacy of Rinvoq in adults with moderate to severe CD and full results from the study will be presented at a future medical meeting and submitted for publication in a peer-reviewed journal.

Recent Events (Continued)

- At the American College of Rheumatology's (ACR) annual meeting, AbbVie shared 38 abstracts from across its rheumatology portfolio that underscored AbbVie's commitment to advancing its portfolio of medicines to help more people living with rheumatic diseases. Highlights included new efficacy data on Rinvoq in people with active PsA and axial involvement, new long-term analysis evaluating the sustainability of response to Rinvoq among patients with RA as well as efficacy and safety data from the KEEPsAKE 1 and KEEPsAKE 2 trials evaluating Skyrizi in adults with PsA treated through 24 weeks.

- AbbVie announced that the FDA granted Breakthrough Therapy Designation (BTD) to investigational telisotuzumab vedotin (Teliso-V) for the treatment of patients with advanced/metastatic epidermal growth factor receptor (EGFR) wild type, nonsquamous non-small cell lung cancer (NSCLC) with high levels of c-Met overexpression whose disease has progressed on or after platinum-based therapy. The BTD is supported by interim data from the ongoing Phase 2 LUMINOSITY study and a Phase 3 study is planned to begin in the first half of 2022.

- At the American Society of Hematology Annual Meeting (ASH), AbbVie presented results from nearly 30 abstracts across 8 types of cancer. Highlights included data from the Phase 2 CAPTIVATE and Phase 3 GLOW studies evaluating minimal residual disease (MRD) and disease-free survival outcomes with fixed duration treatment in patients with chronic lymphocytic leukemia (CLL)/small lymphocytic leukemia (SLL) who received the Imbruvica (ibrutinib) + Venclexta (venetoclax) combination regimen; results from several studies evaluating Venclexta in approved and investigational indications; as well as data evaluating ABBV-383, epcoritamab and lemzoparlimab. Venetoclax is being developed by AbbVie and Roche and is jointly commercialized by AbbVie and Genentech, a member of the Roche Group, in the U.S. and by AbbVie outside of the U.S. Imbruvica is jointly developed and commercialized with Janssen Biotech, Inc. Epcoritamab is being co-developed by Genmab and AbbVie. Lemzoparlimab is being developed through a collaboration with AbbVie and I-Mab.

- Allergan Aesthetics announced the successful completion of its acquisition of Soliton, Inc. The addition of Soliton and its technology complements Allergan Aesthetics' portfolio of non-invasive body contouring treatments to now include a proven treatment for the appearance of cellulite.

- At the American Society for Dermatologic Surgery meeting, Allergan Aesthetics presented 6 abstracts from its leading portfolio of aesthetic treatments and products, which highlighted its approach to innovative science and commitment to bring new and impactful treatments to customers and patients globally. Highlights included two Botox Cosmetic (OnabotulinumtoxinA) abstracts that were recognized as "Best of Cosmetic Oral Abstracts".

- AbbVie announced the FDA approved Vuity (pilocarpine HCl ophthalmic solution) 1.25% for the treatment of presbyopia, commonly known as age-related blurry near vision, in adults. Vuity is the first and only FDA-approved eye drop to treat this common and progressive eye condition that affects nearly half of the U.S. adult population. The approval is supported by two pivotal Phase 3 studies that demonstrated Vuity works in as early as 15 minutes and lasts for up to 6 hours, as measured on day 30, to improve near and intermediate vision without impacting distance vision.

- At the American Academy of Ophthalmology Annual Meeting (AAO), AbbVie presented new data from its leading eye care portfolio. Highlights included new pooled post-hoc analyses and patient-reported outcomes of Vuity 1.25%, analyses on Durysta (bimatoprost intracameral implant) and 3 real-world data studies on the glaucoma patient journey.

- AbbVie announced that it has extended its preclinical oncology research collaboration agreement with the University of Chicago through 2025. Under the agreement, the organizations will continue working together to advance research in several areas, focusing on oncology, and AbbVie gains an option for an exclusive license to certain University of Chicago discoveries made as part of the collaboration.

Full-Year 2022 Outlook

AbbVie is issuing its GAAP diluted EPS guidance for the full-year 2022 of $9.26 to $9.46. AbbVie expects to deliver adjusted diluted EPS for the full-year 2022 of $14.00 to $14.20. The company's 2022 adjusted diluted EPS guidance excludes $4.74 per share of intangible asset amortization expense, non-cash charges for contingent consideration adjustments and other specified items.

About AbbVie

AbbVie's mission is to discover and deliver innovative medicines that solve serious health issues today and address the medical challenges of tomorrow. We strive to have a remarkable impact on people's lives across several key therapeutic areas: immunology, oncology, neuroscience, eye care, virology, women's health and gastroenterology, in addition to products and services across its Allergan Aesthetics portfolio. For more information about AbbVie, please visit us at www.abbvie.com. Follow @abbvie on Twitter, Facebook or LinkedIn.

Conference Call

AbbVie will host an investor conference call today at 8:00 a.m. Central time to discuss our fourth-quarter performance. The call will be webcast through AbbVie's Investor Relations website at investors.abbvie.com. An archived edition of the call will be available after 11:00 a.m. Central time.

Non-GAAP Financial Results

Financial results for 2021 and 2020 are presented on both a reported and a non-GAAP basis. Reported results were prepared in accordance with GAAP and include all revenue and expenses recognized during the period. Non-GAAP results adjust for certain non-cash items and for factors that are unusual or unpredictable, and exclude those costs, expenses, and other specified items presented in the reconciliation tables later in this release. AbbVie's management believes non-GAAP financial measures provide useful information to investors regarding AbbVie's results of operations and assist management, analysts, and investors in evaluating the performance of the business. Non-GAAP financial measures should be considered in addition to, and not as a substitute for, measures of financial performance prepared in accordance with GAAP. The company's 2022 financial guidance is also being provided on both a reported and a non-GAAP basis.

Forward-Looking Statements

Some statements in this news release are, or may be considered, forward-looking statements for purposes of the Private Securities Litigation Reform Act of 1995. The words "believe," "expect," "anticipate," "project" and similar expressions, among others, generally identify forward-looking statements. AbbVie cautions that these forward-looking statements are subject to risks and uncertainties that may cause actual results to differ materially from those indicated in the forward-looking statements. Such risks and uncertainties include, but are not limited to, the failure to realize the expected benefits of AbbVie's acquisition of Allergan or to promptly and effectively integrate Allergan's business, challenges to intellectual property, competition from other products, difficulties inherent in the research and development process, adverse litigation or government action, and changes to laws and regulations applicable to our industry. Additional information about the economic, competitive, governmental, technological and other factors that may affect AbbVie's operations is set forth in Item 1A, "Risk Factors," of AbbVie's 2020 Annual Report on Form 10-K, which has been filed with the Securities and Exchange Commission, as updated by its Quarterly Reports on Form 10-Q and in other documents that AbbVie subsequently files with the Securities and Exchange Commission that update, supplement or supersede such information. AbbVie undertakes no obligation to release publicly any revisions to forward-looking statements as a result of subsequent events or developments, except as required by law.

Media:
Frank Benenati
(224) 668-4169

Investors:
Liz Shea
(847) 935-2211

Todd Bosse
(847) 936-1182

Jeffrey Byrne
(847) 938-2923

AbbVie Inc.
Key Product Revenues
Quarter Ended December 31, 2021
(Unaudited)

	Net Revenues (in millions)			% Change vs. 4Q20				
				Reported			Operational[a]	
	U.S.	Int'l.	Total	U.S.	Int'l.	Total	Int'l.	Total
NET REVENUES	**$11,677**	**$3,209**	**$14,886**	**9.5%**	**0.5%**	**7.4%**	**0.9%**	**7.5%**
Immunology	**5,696**	**1,050**	**6,746**	**14.2**	**8.3**	**13.2**	**9.0**	**13.3**
Humira	4,553	781	5,334	6.0	(9.1)	3.5	(8.8)	3.5
Skyrizi	761	134	895	68.6	82.1	70.5	84.8	70.9
Rinvoq	382	135	517	57.1	>100.0	84.4	>100.0	85.2
Hematologic Oncology	**1,363**	**510**	**1,873**	**(0.7)**	**22.5**	**4.6**	**23.1**	**4.7**
Imbruvica[b]	1,114	271	1,385	(4.3)	4.6	(2.7)	4.6	(2.7)
Venclexta	249	239	488	19.4	51.8	33.3	53.5	34.0
Aesthetics	**877**	**530**	**1,407**	**21.1**	**27.1**	**23.3**	**25.8**	**22.8**
Botox Cosmetic	397	229	626	31.3	20.3	27.0	19.3	26.6
Juvederm Collection	180	252	432	22.8	36.8	30.6	35.3	29.8
Other Aesthetics	300	49	349	9.1	15.3	9.9	13.9	9.7
Neuroscience	**1,440**	**214**	**1,654**	**21.1**	**7.1**	**19.0**	**7.0**	**19.0**
Botox Therapeutic	561	110	671	18.9	15.4	18.3	14.1	18.1
Vraylar	489	—	489	21.8	n/a	21.8	n/a	21.8
Duodopa	29	99	128	0.1	(2.0)	(1.5)	(0.7)	(0.5)
Ubrelvy	183	—	183	>100.0	n/a	>100.0	n/a	>100.0
Other Neuroscience	178	5	183	(19.9)	42.0	(18.9)	34.3	(19.0)
Eye Care	**672**	**288**	**960**	**7.6**	**(4.7)**	**3.6**	**(3.8)**	**3.9**
Lumigan/Ganfort	72	77	149	6.2	(9.2)	(2.4)	(9.0)	(2.3)
Alphagan/Combigan	102	39	141	9.0	(4.3)	4.9	(3.6)	5.1
Restasis	350	14	364	4.9	28.1	5.7	32.8	5.9
Other Eye Care	148	158	306	14.2	(4.8)	3.6	(3.7)	4.2
Women's Health	**216**	**7**	**223**	**(12.5)**	**(18.4)**	**(12.7)**	**(22.1)**	**(12.8)**
Lo Loestrin	123	5	128	(10.6)	16.1	(10.0)	9.9	(10.1)
Orilissa/Oriahnn	37	2	39	3.6	44.1	4.8	37.2	4.6
Other Women's Health	56	—	56	(24.2)	(75.9)	(26.2)	(75.8)	(26.2)
Other Key Products	**1,146**	**283**	**1,429**	**0.9**	**(8.6)**	**(1.1)**	**(6.8)**	**(0.7)**
Mavyret	197	230	427	(10.9)	(11.6)	(11.3)	(9.4)	(10.1)
Creon	327	—	327	7.8	n/a	7.8	n/a	7.8
Lupron	148	44	192	6.1	5.4	6.0	5.5	6.0
Linzess/Constella	278	9	287	0.1	20.4	0.6	17.4	0.5
Synthroid	196	—	196	0.9	n/a	0.9	n/a	0.9

[a] "Operational" comparisons are presented at constant currency rates that reflect comparative local currency net revenues at the prior year's foreign exchange rates.

[b] Reflects profit sharing for Imbruvica international revenues.

n/a = not applicable

AbbVie Inc.
Key Product Revenues
Twelve Months Ended December 31, 2021
(Unaudited)

	Net Revenues (in millions)			% Change vs. 12M20					
				Reported			Comparable Operational[a, b]		
	U.S.	Int'l.	Total	U.S.	Int'l.	Total	U.S.	Int'l.	Total
ADJUSTED NET REVENUES[c]	**$43,435**	**$12,687**	**$56,122**	**24.6%**	**16.1%**	**22.6%**	**12.3%**	**4.7%**	**10.5%**
Immunology	**21,087**	**4,197**	**25,284**	**16.2**	**4.8**	**14.1**	**16.2**	**1.2**	**13.5**
Humira	17,330	3,364	20,694	7.6	(9.6)	4.3	7.6	(12.8)	3.7
Skyrizi	2,486	453	2,939	79.6	>100.0	84.9	79.6	>100.0	84.0
Rinvoq	1,271	380	1,651	94.8	>100.0	>100.0	94.8	>100.0	>100.0
Hematologic Oncology	**5,255**	**1,973**	**7,228**	**2.8**	**28.0**	**8.7**	**2.8**	**26.2**	**8.3**
Imbruvica[d]	4,321	1,087	5,408	0.4	7.7	1.8	0.4	7.7	1.8
Venclexta	934	886	1,820	16.1	66.2	36.1	16.1	60.9	34.0
Aesthetics	**3,350**	**1,883**	**5,233**	**>100.0**	**>100.0**	**>100.0**	**44.7**	**52.2**	**47.3**
Botox Cosmetic*	1,424	808	2,232	>100.0	90.0	>100.0	57.4	42.6	51.8
Juvederm Collection*	658	877	1,535	>100.0	>100.0	>100.0	53.6	61.3	57.9
Other Aesthetics*	1,268	198	1,466	90.2	>100.0	93.0	29.2	56.9	32.1
Neuroscience	**5,061**	**866**	**5,927**	**76.8**	**36.7**	**69.5**	**23.0**	**10.6**	**21.1**
Botox Therapeutic*	2,012	439	2,451	74.3	89.0	76.7	20.5	22.8	20.9
Vraylar*	1,728	—	1,728	81.7	n/a	81.7	24.5	n/a	24.5
Duodopa	102	409	511	(1.0)	4.6	3.4	(1.0)	(0.1)	(0.3)
Ubrelvy*	552	—	552	>100.0	n/a	>100.0	>100.0	n/a	>100.0
Other Neuroscience*	667	18	685	26.3	77.4	27.2	(17.7)	14.2	(17.2)
Eye Care	**2,403**	**1,164**	**3,567**	**65.9**	**58.2**	**63.3**	**5.6**	**2.2**	**4.5**
Lumigan/Ganfort*	273	306	579	64.7	44.1	53.1	(0.1)	(10.2)	(5.6)
Alphagan/Combigan*	373	156	529	66.5	52.5	62.1	5.7	1.7	4.5
Restasis*	1,234	56	1,290	63.3	75.3	63.8	4.1	24.9	4.9
Other Eye Care*	523	646	1,169	72.7	66.1	69.0	12.9	7.6	10.0
Women's Health	**771**	**25**	**796**	**19.1**	**(1.6)**	**18.3**	**(16.0)**	**(33.7)**	**(16.6)**
Lo Loestrin*	423	14	437	21.9	43.3	22.5	(18.5)	(4.9)	(18.2)
Orilissa/Oriahnn	139	6	145	15.4	57.7	16.7	15.4	47.6	16.4
Other Women's Health*	209	5	214	16.2	(57.5)	11.7	(24.8)	(73.9)	(27.7)
Other Key Products	**4,322**	**1,167**	**5,489**	**10.3**	**(3.9)**	**6.9**	**2.8**	**(7.1)**	**0.6**
Mavyret	754	956	1,710	(4.0)	(8.5)	(6.5)	(4.0)	(10.8)	(7.8)
Creon	1,191	—	1,191	6.9	n/a	6.9	6.9	n/a	6.9
Lupron	604	179	783	0.5	18.0	4.0	0.5	15.0	3.4
Linzess/Constella*	1,006	32	1,038	55.1	77.3	55.7	8.0	9.9	8.1
Synthroid	767	—	767	(0.6)	n/a	(0.6)	(0.6)	n/a	(0.6)

[a] "Comparable Operational" comparisons include full-period current year and prior year results for Allergan products, as if the acquisition closed on January 1, 2019, and are presented at constant currency rates that reflect comparative local currency net revenues at the prior year's foreign exchange rates.

[b] All historically reported Allergan revenues have been recast to conform to AbbVie's revenue recognition accounting policies and reporting conventions for certain rebates and discounts. Historically reported Allergan revenues also exclude Zenpep and Viokace product revenues, which were both divested as part of the acquisition, as well as specified items.

[c] Adjusted net revenues exclude specified items. Refer to the Reconciliation of GAAP Reported to Non-GAAP Adjusted Information for further details. Percentage change is calculated using adjusted net revenues.

[d] Reflects profit sharing for Imbruvica international revenues.

* Represents product(s) acquired as part of the Allergan acquisition.

n/a = not applicable

AbbVie Inc.
Consolidated Statements of Earnings
Quarter and Twelve Months Ended December 31, 2021 and 2020
(Unaudited) (In millions, except per share data)

	Fourth Quarter Ended December 31		Twelve Months Ended December 31	
	2021	2020	2021	2020
Net revenues	$ 14,886	$ 13,858	$ 56,197	$ 45,804
Cost of products sold	4,320	4,684	17,446	15,387
Selling, general and administrative	3,260	3,231	12,349	11,299
Research and development	1,827	1,890	7,084	6,557
Acquired in-process research and development	405	300	962	1,198
Other operating expense, net	—	—	432	—
Total operating costs and expenses	9,812	10,105	38,273	34,441
Operating earnings	5,074	3,753	17,924	11,363
Interest expense, net	571	618	2,384	2,280
Net foreign exchange loss	16	17	51	71
Other expense, net	216	4,625	2,500	5,614
Earnings (loss) before income tax expense	4,271	(1,507)	12,989	3,398
Income tax expense (benefit)	226	(1,545)	1,440	(1,224)
Net earnings	4,045	38	11,549	4,622
Net earnings attributable to noncontrolling interest	1	2	7	6
Net earnings attributable to AbbVie Inc.	$ 4,044	$ 36	$ 11,542	$ 4,616
Diluted earnings per share attributable to AbbVie Inc.	$ 2.26	$ 0.01	$ 6.45	$ 2.72
Adjusted diluted earnings per share[a]	$ 3.31	$ 2.92	$ 12.70	$ 10.56
Weighted-average diluted shares outstanding	1,778	1,776	1,777	1,673

[a] Refer to the Reconciliation of GAAP Reported to Non-GAAP Adjusted Information for further details. Weighted-average diluted shares outstanding includes the effect of dilutive securities.

AbbVie Inc.
Reconciliation of GAAP Reported to Non-GAAP Adjusted Information
Quarter Ended December 31, 2021
(Unaudited) (In millions, except per share data)

1. Specified items impacted results as follows:

	4Q21		
	Earnings		Diluted
	Pre-tax	After-tax[a]	EPS
As reported (GAAP)	**$ 4,271**	**$ 4,044**	**$ 2.26**
Adjusted for specified items:			
Intangible asset amortization	1,806	1,490	0.84
Acquisition and integration costs	(191)	(212)	(0.12)
Acquired IPR&D	405	405	0.23
Change in fair value of contingent consideration	232	232	0.13
Litigation matters	200	167	0.09
Impacts related to tax law changes	—	(265)	(0.15)
Other	41	58	0.03
As adjusted (non-GAAP)	**$ 6,764**	**$ 5,919**	**$ 3.31**

[a] Represents net earnings attributable to AbbVie Inc.

Acquisition and integration costs reflect a recovery of certain Allergan acquisition-related regulatory fees partially offset by Allergan-related integration costs and Soliton acquisition costs. Acquired IPR&D represents initial costs to acquire rights to in-process R&D projects through R&D collaborations, licensing arrangements or other asset acquisitions. Other primarily includes COVID-19 related expenses and tax related items.

2. The impact of the specified items by line item was as follows:

	4Q21				
	Cost of products sold	SG&A	R&D	Acquired IPR&D	Other expense, net
As reported (GAAP)	**$ 4,320**	**$ 3,260**	**$ 1,827**	**$ 405**	**$ 216**
Adjusted for specified items:					
Intangible asset amortization	(1,806)	—	—	—	—
Acquisition and integration costs	(43)	250	(16)	—	—
Acquired IPR&D	—	—	—	(405)	—
Change in fair value of contingent consideration	—	—	—	—	(232)
Litigation matters	—	(200)	—	—	—
Other	(23)	(3)	(13)	—	(2)
As adjusted (non-GAAP)	**$ 2,448**	**$ 3,307**	**$ 1,798**	**$ —**	**$ (18)**

3. The adjusted tax rate for the fourth quarter of 2021 was 12.5 percent, as detailed below:

	4Q21		
	Pre-tax earnings	Income taxes	Tax rate
As reported (GAAP)	**$ 4,271**	**$ 226**	**5.3 %**
Specified items	2,493	618	24.8 %
As adjusted (non-GAAP)	**$ 6,764**	**$ 844**	**12.5 %**

AbbVie Inc.
Reconciliation of GAAP Reported to Non-GAAP Adjusted Information
Quarter Ended December 31, 2020
(Unaudited) (In millions, except per share data)

1. Specified items impacted results as follows:

	4Q20		
	Earnings (Loss)		Diluted
	Pre-tax	After-tax[a]	EPS
As reported (GAAP)	**$ (1,507)**	**$ 36**	**$ 0.01**
Adjusted for specified items:			
Intangible asset amortization	1,838	1,444	0.81
Acquisition and integration costs	467	399	0.22
Milestones and other R&D expenses	48	39	0.02
Acquired IPR&D	300	296	0.16
Change in fair value of contingent consideration	4,675	4,671	2.63
Tax audit settlements	—	(140)	(0.08)
Impacts related to tax law changes	—	(1,492)	(0.84)
Other	92	(28)	(0.01)
As adjusted (non-GAAP)	**$ 5,913**	**$ 5,225**	**$ 2.92**

[a] Represents net earnings attributable to AbbVie Inc.

Acquisition and integration costs reflect integration costs and amortization of the acquisition date fair value step-up for inventory related to the Allergan acquisition. Milestones and other R&D expenses include milestone payments for previously announced collaborations. Acquired IPR&D represents initial costs to acquire rights to in-process R&D projects through R&D collaborations, licensing arrangements or other asset acquisitions. Other primarily includes tax related items and COVID-19 related expenses.

2. The impact of the specified items by line item was as follows:

	4Q20				
	Cost of products sold	SG&A	R&D	Acquired IPR&D	Other expense, net
As reported (GAAP)	**$ 4,684**	**$ 3,231**	**$ 1,890**	**$ 300**	**$ 4,625**
Adjusted for specified items:					
Intangible asset amortization	(1,838)	—	—	—	—
Acquisition and integration costs	(272)	(126)	(69)	—	—
Milestones and other R&D expenses	—	—	(48)	—	—
Acquired IPR&D	—	—	—	(300)	—
Change in fair value of contingent consideration	—	—	—	—	(4,675)
Other	(51)	(16)	(22)	—	(3)
As adjusted (non-GAAP)	**$ 2,523**	**$ 3,089**	**$ 1,751**	**$ —**	**$ (53)**

3. The adjusted tax rate for the fourth quarter of 2020 was 11.6 percent, as detailed below:

	4Q20		
	Pre-tax earnings (loss)	Income taxes	Tax rate
As reported (GAAP)	**$ (1,507)**	**$ (1,545)**	**102.5 %**
Specified items	7,420	2,231	30.1 %
As adjusted (non-GAAP)	**$ 5,913**	**$ 686**	**11.6 %**

AbbVie Inc.
Reconciliation of GAAP Reported to Non-GAAP Adjusted Information
Twelve Months Ended December 31, 2021
(Unaudited) (In millions, except per share data)

1. Specified items impacted results as follows:

	12M21		
	Earnings		Diluted
	Pre-tax	After-tax[a]	EPS
As reported (GAAP)	**$ 12,989**	**$ 11,542**	**$ 6.45**
Adjusted for specified items:			
Intangible asset amortization	7,718	6,419	3.60
Acquisition and integration costs	344	215	0.12
Milestones and other R&D expenses	359	307	0.17
Acquired IPR&D	962	948	0.53
Calico collaboration	500	500	0.28
Change in fair value of contingent consideration	2,679	2,677	1.50
Litigation matters	307	253	0.14
Impacts related to tax law changes	—	(265)	(0.15)
Other	88	100	0.06
As adjusted (non-GAAP)	**$ 25,946**	**$ 22,696**	**$ 12.70**

[a] Represents net earnings attributable to AbbVie Inc.

Acquisition and integration costs reflect Allergan integration costs, Soliton acquisition costs as well as amortization of the acquisition date fair value step-up for inventory related to the Allergan acquisition partially offset by a recovery of certain Allergan acquisition-related regulatory fees. Milestones and other R&D expenses include milestone payments for previously announced collaborations and the purchase of FDA priority review vouchers from third parties. Acquired IPR&D represents initial costs to acquire rights to in-process R&D projects through R&D collaborations, licensing arrangements or other asset acquisitions. Other primarily includes COVID-19 related expenses, restructuring charges associated with streamlining global operations and tax related items, offset by milestone revenue under an existing collaboration agreement.

2. The impact of the specified items by line item was as follows:

	12M21						
	Net revenues	Cost of products sold	SG&A	R&D	Acquired IPR&D	Other operating expense, net	Other expense, net
As reported (GAAP)	**$ 56,197**	**$ 17,446**	**$ 12,349**	**$ 7,084**	**$ 962**	**$ 432**	**$ 2,500**
Adjusted for specified items:							
Intangible asset amortization	—	(7,718)	—	—	—	—	—
Acquisition and integration costs	—	(215)	(25)	(104)	—	—	—
Milestones and other R&D expenses	—	—	—	(359)	—	—	—
Acquired IPR&D	—	—	—	—	(962)	—	—
Calico collaboration	—	—	—	—	—	(500)	—
Change in fair value of contingent consideration	—	—	—	—	—	—	(2,679)
Litigation matters	—	—	(307)	—	—	—	—
Other	(75)	(88)	(53)	(103)	—	68	13
As adjusted (non-GAAP)	**$ 56,122**	**$ 9,425**	**$ 11,964**	**$ 6,518**	**$ —**	**$ —**	**$ (166)**

3. The adjusted tax rate for the full-year 2021 was 12.5 percent, as detailed below:

	12M21		
	Pre-tax earnings	Income taxes	Tax rate
As reported (GAAP)	**$ 12,989**	**$ 1,440**	**11.1 %**
Specified items	12,957	1,803	13.9 %
As adjusted (non-GAAP)	**$ 25,946**	**$ 3,243**	**12.5 %**

AbbVie Inc.
Reconciliation of GAAP Reported to Non-GAAP Adjusted Information
Twelve Months Ended December 31, 2020
(Unaudited) (In millions, except per share data)

1. Specified items impacted results as follows:

	12M20		
	Earnings		Diluted
	Pre-tax	After-tax[a]	EPS
As reported (GAAP)	**$ 3,398**	**$ 4,616**	**$ 2.72**
Adjusted for specified items:			
Intangible asset amortization	5,805	4,805	2.87
Acquisition and integration costs	3,366	3,023	1.81
Milestones and other R&D expenses	273	241	0.14
Acquired IPR&D	1,198	1,194	0.71
Change in fair value of contingent consideration	5,753	5,749	3.43
Tax audit settlements	—	(200)	(0.12)
Impacts related to tax law changes	—	(1,689)	(1.02)
Other	239	42	0.02
As adjusted (non-GAAP)	**$ 20,032**	**$ 17,781**	**$ 10.56**

[a] Represents net earnings attributable to AbbVie Inc.

Acquisition and integration costs reflect transaction and financing costs, compensation expense and other integration costs as well as amortization of the acquisition date fair value step-up for inventory related to the Allergan acquisition. Milestones and other R&D expenses include milestone payments for previously announced collaborations and the purchase of an FDA priority review voucher from a third party. Acquired IPR&D represents initial costs to acquire rights to in-process R&D projects through R&D collaborations, licensing arrangements or other asset acquisitions. Other primarily includes tax related items and COVID-19 related charitable contributions and expenses.

2. The impact of the specified items by line item was as follows:

	12M20							
	Net revenues	Cost of products sold	SG&A	R&D	Acquired IPR&D	Interest expense, net	Net foreign exchange loss	Other expense, net
As reported (GAAP)	**$ 45,804**	**$ 15,387**	**$ 11,299**	**$ 6,557**	**$ 1,198**	**$ 2,280**	**$ 71**	**$ 5,614**
Adjusted for specified items:								
Intangible asset amortization	—	(5,805)	—	—	—	—	—	—
Acquisition and integration costs	—	(1,292)	(1,416)	(384)	—	(274)	—	—
Milestones and other R&D expenses	—	—	—	(273)	—	—	—	—
Acquired IPR&D	—	—	—	—	(1,198)	—	—	—
Change in fair value of contingent consideration	—	—	—	—	—	—	—	(5,753)
Other	(20)	(115)	(80)	(70)	—	—	9	(3)
As adjusted (non-GAAP)	**$ 45,784**	**$ 8,175**	**$ 9,803**	**$ 5,830**	**$ —**	**$ 2,006**	**$ 80**	**$ (142)**

3. The adjusted tax rate for the full-year 2020 was 11.2 percent, as detailed below:

	12M20		
	Pre-tax earnings	Income taxes	Tax rate
As reported (GAAP)	**$ 3,398**	**$ (1,224)**	**(36.0)%**
Specified items	16,634	3,469	20.9 %
As adjusted (non-GAAP)	**$ 20,032**	**$ 2,245**	**11.2 %**

ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 151−157

Journal of Clinical Epidemiology

GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes

Gordon Guyatt[a,j,*], Andrew D. Oxman[b], Shahnaz Sultan[c], Jan Brozek[d], Paul Glasziou[e], Pablo Alonso-Coello[f], David Atkins[g], Regina Kunz[h,i], Victor Montori[j], Roman Jaeschke[k], David Rind[l], Philipp Dahm[m], Elie A. Akl[n], Joerg Meerpohl[o,p], Gunn Vist[b], Elise Berliner[q], Susan Norris[r], Yngve Falck-Ytter[r], Holger J. Schünemann[a]

[a]*Departments of Clinical Epidemiology and Biostatistics and Medicine, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[b]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St. Olavs plass, 0130 Oslo, Norway*
[c]*Department of Medicine, Division of Gastroenterology, Hepatology, and Nutrition, University of Florida, Gainesville, FL, USA*
[d]*University of Oxford, Oxford, United Kingdom*
[e]*Iberoamerican Cochrane Centre, CIBERESP-IIB Sant Pau, Barcelona 08041, Spain*
[f]*QUERI Program, Office of Research and Development, Department of Veterans Affairs, Washington, DC, USA*
[g]*Academy of Swiss Insurance Medicine (asim) University Hospital Basel Petergraben 4 CH-4031, Basel, Switzerland*
[h]*The Basel Institute of Clinical Epidemiology, University Hospital Basel Hebelstrasse 10, 4031 Basel, Switzerland*
[i]*Knowledge and Evaluation Research Unit, Mayo Clinic, Rochester, MN, USA*
[j]*Department of Medicine, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[k]*Harvard Medical School, UpToDate, Boston, USA*
[l]*Department of Urology, University of Florida, Gainesville, FL, USA*
[m]*Department of Medicine, State University of New York at Buffalo, Buffalo, NY, USA*
[n]*German Cochrane Center, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, 79104 Freiburg, Germany*
[o]*Division of Pediatric Hematology and Oncology, Department of Pediatric and Adolescent Medicine, University Medical Center Freiburg, 79106 Freiburg, Germany*
[p]*Technology Assessment Program, Center for Outcomes and Evidence, Agency for Healthcare Research and Quality, 540 Gaither Road, Rockville, MD 20850, USA*
[q]*Oregon Health and Science University, Department of Medical Informatics and Clinical Epidemiology, Portland, OR 97239-3098, USA*
[r]*Division of Gastroenterology, Case and VA Medical Center, Case Western Reserve University, Cleveland, OH 44106, USA*

Accepted 15 January 2012; Published online 27 April 2012

Abstract

GRADE requires guideline developers to make an overall rating of confidence in estimates of effect (quality of evidence—high, moderate, low, or very low) for each important or critical outcome. GRADE suggests, for each outcome, the initial separate consideration of five domains of reasons for rating down the confidence in effect estimates, thereby allowing systematic review authors and guideline developers to arrive at an outcome-specific rating of confidence. Although this rating system represents discrete steps on an ordinal scale, it is helpful to view confidence in estimates as a continuum, and the final rating of confidence may differ from that suggested by separate consideration of each domain.

An overall rating of confidence in estimates of effect is only relevant in settings when recommendations are being made. In general, it is based on the critical outcome that provides the lowest confidence.

Keywords: GRADE; Quality of evidence; Confidence in estimates; Guideline methodology; Systematic review methodology; Values and preferences

1. Introduction

In prior studies in this series devoted to exploring GRADE's approach to rating confidence in estimates of effect (quality of evidence) and grading strength of recommendations (guidance for practice) we have dealt with issues of framing the question [1]; introduced GRADE's

The GRADE system has been developed by the GRADE Working Group. The named authors drafted and revised this article. A complete list of contributors to this series can be found on the JCE Web site.

* Corresponding author. Gordon H. Guyatt, CLARITY Research Group, Department of Clinical Epidemiology & Biostatistics, Room 2C12, 1200 Main Street West Hamilton, ON, L8N 3Z5, Canada. Tel.: +905-525-9140; fax: +905-523-8781.

E-mail address: guyatt@mcmaster.ca (G. Guyatt).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2012.01.006

What is new?

Key points

GRADE requires a rating of confidence in effect estimates (quality of evidence) for each outcome.

Rating of confidence of evidence requires a gestalt that simultaneously considers all eight domains (risk of bias, precision, consistency, and so forth)

Guideline developers using GRADE will subsequently make an overall rating of confidence in effect estimates across all outcomes based on those outcomes they consider critical to their recommendation.

Optimal application of GRADE requires making the reasons for key judgments transparent.

conceptual approach to rating the confidence in a body of evidence [2]; and presented five reasons for rating down the confidence in effect estimates (risk of bias [3], imprecision [4], inconsistency [5], indirectness [6], and publication bias [7]) and three reasons for rating up the confidence in effect estimates [8] (a large magnitude of effect, a dose-response gradient, and a situation in which plausible biases, if present, would serve to increase our confidence in the effect estimate), as well as dealing with issues specific to resource use. This 11th article in the series will focus on (1) summarizing the confidence in effect estimates across a single outcome for each important or critical outcome and (2) determining the confidence in effect estimates across all critical outcomes.

2. Summarizing the confidence in effect estimates for individual outcomes

GRADE's approach to rating down (or not) with respect to each of five criteria and to rating up (or not) with respect to three others is sometimes straightforward and enhances the transparency of the system. Most commonly, authors will be comfortable with the rating of confidence in estimate of effect that results from considering each criterion separately. Not infrequently, however, if ratings are applied in a blanket or rote fashion without considering context and the relation of one criterion to another, the confidence rating could be problematic. Specifically, ratings of individual domains could result in an overall rating of confidence in effect estimates on a particular outcome that does not correspond well to an integrated assessment or the gestalt of confidence in estimates of effect. In such instances, an adjustment in the final rating based on that gestalt is required.

Consider a systematic review of randomized trials of flavonoids for the treatment of hemorrhoids that produced a pooled estimate of a relative risk of persisting symptoms (lack of improvement) of 0.42 (95% confidence interval [CI] 0.28−0.61) [9]. Table 1 presents an evidence profile summarizing the evidence regarding two outcomes: persisting symptoms and adverse effects of the intervention. The profile presents the number of studies and patients, considerations related to the five possible reasons for rating down confidence in effect estimates (summarized in the table with expansions in the associated footnotes), and the best estimates and CIs around relative and absolute effects.

Consider now the possible reasons for rating down confidence in effect estimates. In most studies, the published articles left uncertainty whether allocation was concealed (though blinding in most suggests the likelihood of concealment), and all studies used unvalidated measures of symptoms. Given these limitations, one could reasonably argue either for or against rating down for risk of bias.

Fig. 1 presents a forest plot depicting the results of the review. The point estimates from individual studies are quite variable, and some of the CIs overlap little. The test for heterogeneity is highly significant and the I^2 large. All these observations suggest rating down for inconsistency among studies. On the other hand, all point estimates suggest benefit, and one might argue that it is inappropriate to rate down for inconsistency when the only uncertainty appears to be whether the magnitude of the treatment effect is moderate or very large. For instance, if undesirable consequences of an intervention are minimal, even a modest treatment effect may warrant a strong recommendation in favor of that treatment. If, in such a circumstance, the basis of doubt is whether the true effect is modest or large, rating down for inconsistency may well be inappropriate.

All available randomized trials were of small or moderate size (from 40 to 234 patients), and all were industry funded. This is a situation that raises the possibility of publication bias. In addition, one could interpret the funnel plot as suggesting the possibility of publication bias, with three small, very positive studies and no corresponding studies with small or negligible effects (Fig. 2). This line of reasoning would suggest rating down confidence in the estimate for publication bias. On the other hand, the number of studies is insufficient to meet rigorous criteria for creating a funnel plot [10] and one could argue that the case for publication bias is speculative in which case one would not rate down.

Thus, for three of the five domains in which one might rate down confidence in effect estimates (risk of bias, inconsistency, and publication bias) one could reasonably make the case for rating down or for not doing so. The situation is further complicated by the magnitude of effect: the relative risk of persisting symptoms (0.41) is slightly less than 0.5, raising the possibility of rating confidence up for the magnitude of effect. A generous reviewer, who in each case is inclined to view the results favorably, would interpret the body of evidence from these flavonoid studies as high quality (i.e., would not rate down the quality). A less generous reviewer, who decides to rate down the

Table 1. GRADE Evidence Profile: flavonoids for patients with symptomatic hemorrhoids (question: flavonoids for patients with symptomatic hemorrhoids?; setting: outpatients)

Quality assessment						Summary of findings					
						No. of patients			Absolute risk		
No of studies (design)	Risk of bias	Inconsistency	Indirectness	Imprecision	Publication bias	No treatment	Flavonoids	Relative risk (95% CI)	Control rate	Risk difference (95% CI)	Quality
Persisting symptoms/lack of improvement											
Nine (RCT)	Concealment not clear in most studies Outcome measures not validated[a]	P-value on test for heterogeneity < 0.0001 I^2 70.4[b]	No serious indirectness	No serious imprecision	All studies industry funded?[c]	218/384	93/398	RR 0.41 (0.27–0.62)	551/1,000	226 fewer per 1,000 (149–342)	Moderate quality because of publication bias[d]
Adverse effects											
13 (RCT)	Lack of concealment and unvalidated questionnaires[a]	No serious inconsistency	No serious indirectness	CI includes reduction to doubling of adverse effects[e]	All studies industry funded[c]	20/681	28/704	RR 1.22 (0.69–2.15)	60/1,000	Not significant	Low quality because of publication bias and imprecision

Abbreviations: CI, confidence interval; RR, relative risk; RCT, randomized controlled trial.

The table highlights the three questionable criteria in which reviewers might either rate down or not—study limitations, inconsistency, and publication bias—and how the final judgments could vary if one came to positive judgments on all three (e.g., for persisting symptoms, high-quality evidence) or negative judgments on all three (e.g., for persisting symptoms, very low-quality evidence).

[a] Allocation concealment unclear in most studies though blinding suggests the likelihood of concealment in most. The outcomes summarized here was failure to improve symptoms and side effects. These were measured by unvalidated questionnaires in each study. The questions, however, were simple and straightforward, final decision was not to rate down for risk of bias.

[b] Although the I^2 is large and the test for heterogeneity very highly significant, all studies but one suggest benefit, and uncertainty appears to be the magnitude of effect rather than whether there is an effect. Final decision not to rate down for inconsistency.

[c] Not only are all studies industry funded, but they are all of small or moderate size. Furthermore, the funnel plot (Fig. 2) could be interpreted as suggesting the possibility of publication bias. Final decision: rate down for likelihood of publication bias.

[d] We rated down for publication bias. Although there also was concern about a high risk of bias and inconsistency, we did not further rate down the quality of evidence because not every criterion appeared to justify rating down by one level.

[e] The lower boundary of the CI would suggest no treatment-induced adverse effects, whereas the upper boundary suggests more than a doubling of adverse effects relative to placebo.

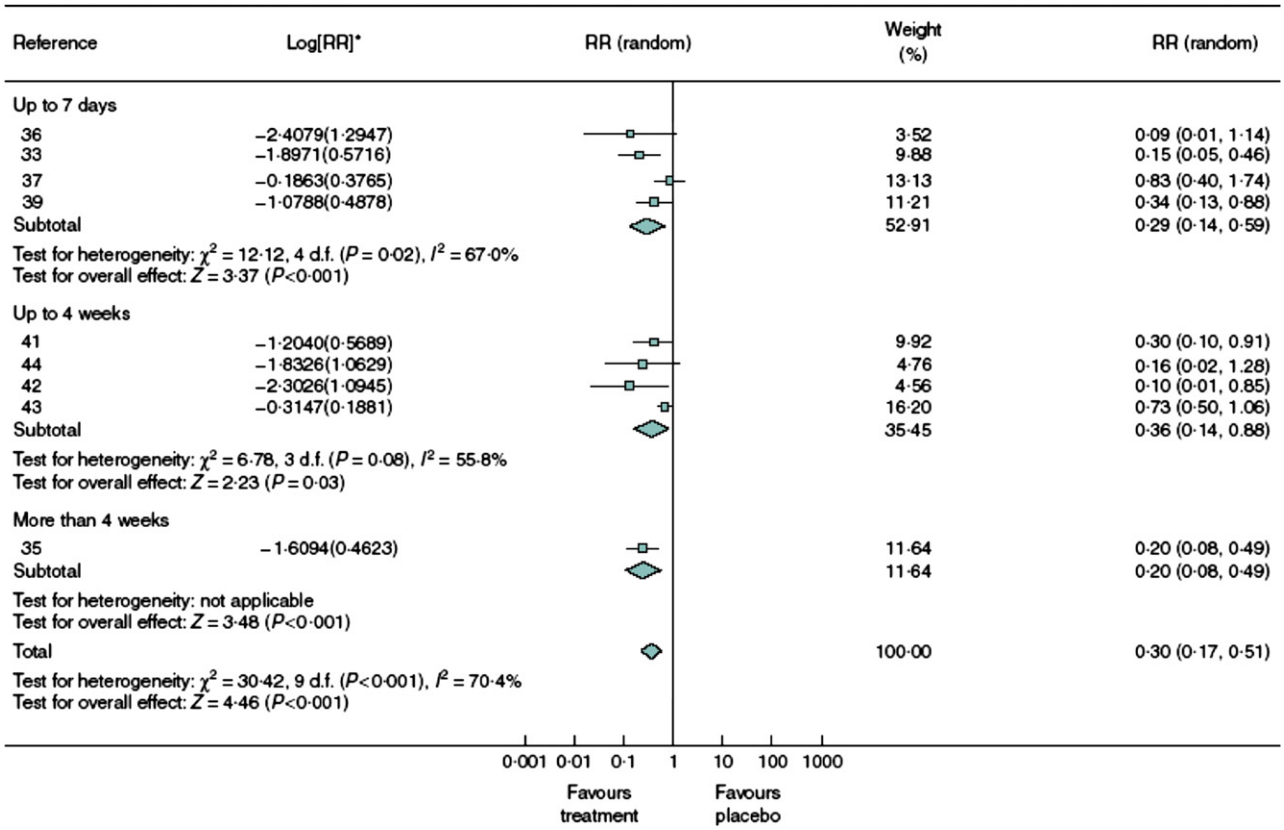


Fig. 1. Forest plot of the results of a systematic review of flavonoids for the treatment of hemorrhoids for the outcome of persisting symptoms or lack of improvement.

evidence in each case and rejects rating up for magnitude of effect, would judge the evidence warranting very low confidence. Both reviewers, having made judgments for individual criteria, might be dismayed that the overall rating (high or very low) does not really capture their confidence in effect estimates.

This example highlights the fact that each criterion for rating quality of evidence up or down reflects not discrete categories but a continuum from minimal limitations to very serious limitations. When the body of evidence is intermediate with respect to a particular criterion, the decision whether a study falls above or below the threshold for rating confidence up or down (by one or two levels) may be arbitrary. In such instances, it is particularly desirable to describe the rationale for the final decisions.

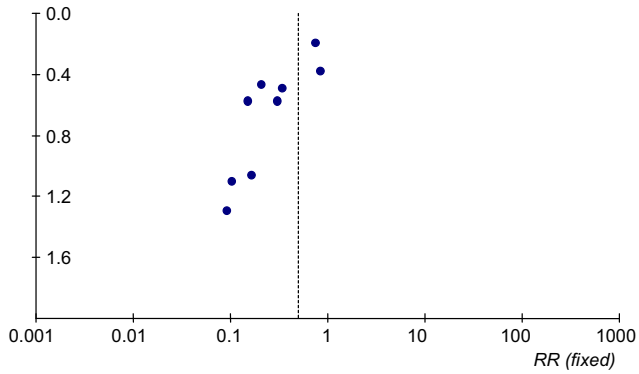


Fig. 2. Funnel plot of studies of flavonoids for ameliorating symptoms in patients with hemorrhoids.

In the case of flavonoids for hemorrhoids, both reviewers—charitable and harsh with respect to individual domains—may, taking a broad look at the evidence, agree that overall it lies on the border of moderate to low quality evidence (which was the conclusion of the authors of the review) [9]. In that case, reviewers may pick one or two domains (risk of bias, inconsistency, or publication bias) of limitations that would explain their reasoning. For example, the associated explanation could read: "We rated down for publication bias. Although there was also concern about a high risk of bias and inconsistency, we did not further rate down confidence in effect estimates because not every criterion appeared to justify rating down by one level." This reflects the necessity to take an overall or gestalt view of the body of evidence. In the evidence profile presentation (Table 1), the final decision is that the body of evidence warrants of moderate confidence, and the chosen reason for rating down confidence is likely publication bias.

Having difficulties about placing the evidence in either the moderate or low confidence category emphasizes that

the overall confidence rating also is a continuum, and contextual decisions are necessary when confidence is near the threshold between categories. The authors of the review acknowledged this by suggesting that ratings of either moderate or low confidence would be reasonable.

We encourage review and guideline authors to be explicit when they encounter similar situations, acknowledging borderline decisions in one or more domains. The evidence profile (Table 1) demonstrates such a presentation (note in particular footnote d).

Despite the limitations of breaking continua into discrete categories, treating each domain for rating confidence up or down as a discrete category enhances transparency. Indeed, the example highlights once again that the great merit of GRADE is not that it necessarily ensures reproducible judgments (observers will inevitably differ in close-call situations when rating up or down for individual domains or for the overall confidence per outcome) but that it achieves explicit and transparent judgment. In such close-call situations, apparent disagreement about whether to rate confidence up or down may represent very little disagreement on a continuum if that disagreement occurs near a threshold between categories (i.e., the threshold between rating down and not rating down). Furthermore, when the overall confidence is near a threshold (e.g., moderate or low confidence), systematic reviewers and guideline developers using GRADE may reduce their angst by recognizing that the disagreement, when the confidence rating is viewed as a continuum, is small.

3. Determining the confidence in effect estimates across outcomes

GRADE is the first formal system of rating quality of evidence to acknowledge that quality may differ across outcomes and to explicitly address this issue. For systematic reviews that are not associated with recommendations, and therefore do not require an overall confidence rating across outcomes, we suggest presenting confidence ratings for each important outcome and not determining the confidence in effect estimates across outcomes.

Such systematic reviews may, however, subsequently inform guidelines that do require implicit or explicit judgments about the overall confidence in effect estimates. It is better to be explicit, and it is logical that the overall confidence in effect estimates cannot be higher than the lowest confidence in effect estimates for any outcome that is critical for a decision. We therefore suggest applying the lowest confidence rating of the critical outcomes as the overall confidence associated with a recommendation. This requires distinguishing between outcomes that are critical and ones that are important but not critical.

Consider a systematic review of alternative strategies for Whipple resection for pancreatic cancer, one of which preserves the pylorus and the other, the standard approach, which does not [11]. The evidence in this review for different outcomes varied from moderate to very low confidence in effect estimates (Table 2). In cases such as this, guideline developers must consider whether undesirable consequences of therapy are important but not critical to the decision regarding the optimal management strategy or whether they are critical. If an outcome for which evidence is of lower quality is a critical outcome for decision making, then the rating of overall quality of the evidence must reflect this lower quality evidence. If the outcome for which confidence is lower is an important but not critical outcome, the overall rating will reflect the higher confidence in estimates from the critical outcomes.

Thus, for this example, if those making recommendations felt that gastric emptying problems were critical, the overall rating of the confidence in effect estimates would be very low. If gastric emptying were important but not critical, the overall confidence would be low (on the basis of results from the clearly critical perioperative mortality) despite the presence of moderate confidence regarding 5-year survival.

4. Which outcomes are critical may depend on the evidence

The overall confidence in effect estimates may not come from the outcomes judged critical at the beginning of the guideline development process—that is, judgments about what is critical may change when considering the results. For instance, a particular adverse event (e.g., severe nausea and vomiting) may be considered critical at the outset. If it turns out, however, that the event occurs very infrequently—say, less than 3% of patients—the final decision may be that the adverse effect is important but not critical.

Consider, once again, the flavonoids for hemorrhoids review (Table 1) [9]. In addition to the risk of bias (concealment not explicit, questionnaires not validated) and publication bias problems associated with the primary outcome of persisting symptoms, the adverse effect outcome suffers from imprecision. Therefore, whatever judgment of confidence one might make about persisting symptoms, adverse effects would warrant lower confidence. However, even assuming the boundary of the CI associated with the largest increase in adverse effects (an approximate doubling in comparison to placebo) represented the true impact of treatment, the adverse effects would still be relatively infrequent (approximately 6.3%) and minor in nature. Despite these considerations, some might consider the adverse effects critical and thus rate the overall confidence in effect estimates low. Others would not and may therefore rate the overall confidence in effect as moderate.

Consider the choice facing individuals without documented coronary heart disease (CHD) but at high risk (e.g., male smokers over 60 with hypertension, elevated

Table 2. GRADE Evidence Profile: different resection strategies for pancreatic carcinoma associated with different evidence quality of different outcomes (question: pylorus-preserving pancreaticoduodenectomy vs. standard Whipple pancreaticoduodenectomy in pancreatic or periampullary cancer?; setting: inpatients)

Quality assessment						Summary of findings					
						No. of patients			Absolute effect		
No of studies (design)	Limitations	Inconsistency	Indirectness	Imprecision	Publication bias	SWPD	PPPD	RR[a] (95% CI)	Control rate	Risk difference (95% CI)	Quality
Mortality at 5 years											
Three (RCT)	Serious limitations[b]	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	94/114	93/115	RR 0.98 (0.87–1.11)	825/1,000	20 fewer per 1,000 (−120 to +80)	⊕⊕⊕○ Moderate
In-hospital mortality											
Six (RCT)	Serious limitations[b]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	Undetected	12/244	4/246	RR 0.40 (0.14–1.13)	49/1,000	20 fewer per 1,000 (−50 to +10)	⊕⊕○○ Low
Biliary leaks											
Three (RCT)	Serious limitations[b]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	Undetected	0/133	2/135	RR 4.77 (0.23–97.96)	0/1,000	20 more per 1,000 (−20 to +50)	⊕⊕○○ Low
Delayed gastric emptying											
Five (RCT)	Serious limitations[b]	Serious inconsistency[d]	No serious indirectness	Serious imprecision[c]	Undetected	56/220	66/222	RR 1.52 (0.74–3.14)	255/1,000	110 more per 1,000 (−80 to +290)	⊕○○○ Very low
Blood transfusions (units)[e]									Best estimate SWPD group	WMD (95% CI)	
Five (RCT)	Serious limitations[b]	No serious inconsistency	No serious indirectness	No serious imprecision	Undetected	320		—[e]	2.45	WMD −0.66 (−1.16 to −0.25)	⊕⊕⊕○ Moderate
Hospital stay (days)[e]											
Five(RCT)	Serious limitations[b]	No serious inconsistency	No serious indirectness	Serious imprecision[c]	Undetected	446		—[e]	19.17	WMD −1.45 (−3.28 to +0.38)	⊕⊕○○ Low

Abbreviations: SWPD, standard Whipple pancreaticoduodenectomy; PPPD, Pylorus-preserving pancreaticoduodenectomy; WMD, weighted mean difference.

[a] All data based on random effect models.

[b] Unclear allocation concealment in all studies, patients blinded in only one study, outcome assessors not blinded in any study, >20% loss to follow-up in three studies, not analyzed using intention to treat in one study.

[c] CI includes possible benefit from both surgical approaches.

[d] Unexplained heterogeneity; $I^2 = 72.6\%$, $P = 0.006$.

[e] Continuous outcome, therefore no relative effect is given.

cholesterol despite attempts at reduction with diet, diabetes, and a family history of CHD): should they use statins to lower their risk of cardiovascular events? A meta-analysis of rigorous randomized trials in such individuals demonstrated consistent, statistically significant reductions in major CHD events and stroke but nonsignificant reductions in CHD deaths [12]. Serious adverse effects were unusual, and all adverse effects were readily reversible with drug discontinuation [13].

Guideline developers considering a recommendation for or against statins in high-risk individuals are likely to start the process of arriving at a recommendation considering all four outcomes (i.e., death from cardiovascular causes, myocardial infarction, stroke, and adverse effects) as critical. In reviewing the evidence, they find that for three of the four outcomes (myocardial infarction, stroke, and toxicity) the evidence warrants high confidence. For CHD deaths, however, because of imprecision, evidence warrants moderate confidence. Should the overall confidence rating across outcomes be high or moderate?

The judgments made at the beginning of the review process suggest that the answer is ''moderate.'' Most patients, however, once it is established that their risk of stroke and major coronary events decreases with statins, would find compelling reason to use the medication. Whether CHD mortality decreases is (as long as it is very unlikely it increases) no longer relevant to the decision. Considering this, the overall confidence rating is most appropriately designated as high confidence.

The principle is that if there is higher confidence in some critical outcomes to support a decision in favor of an intervention (i.e., benefits on critical outcomes clearly outweigh undesirable effects of the intervention, for which there also is high-quality evidence) one need not rate down confidence because of lower confidence in other critical outcomes that support the same recommendation. To put it another way: an outcome is no longer critical if, across the range of possible effect of the intervention on that outcome, the recommendation or its strength would remain unchanged. Such judgments require careful consideration and are probably rare.

5. Conclusions

GRADE defines criteria for rating the confidence in effect estimates for a given outcome, thereby allowing systematic review authors and guideline developers to arrive at an outcome-specific confidence in effect estimates rating. Although this rating system represents discrete steps on an ordinal scale, it is helpful to view confidence in effect estimates as a continuum. An overall confidence in effect estimates rating across outcomes is only relevant in settings when recommendations are being made. In general, it is based on the critical outcome that provides the lowest confidence in effect estimates.

References

[1] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395−400.

[2] Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401−6.

[3] Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence−study limitations (risk of bias). J Clin Epidemiol 2011;64:407−15.

[4] Guyatt G, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence-imprecision. J Clin Epidemiol 2011;64:1283−93.

[5] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence-inconsistency. J Clin Epidemiol 2011;64:1294−302.

[6] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence-indirectness. J Clin Epidemiol 2011;64:1303−10.

[7] Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence-publication bias. J Clin Epidemiol 2011;64:1277−82.

[8] Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol 2011;64:1311−6.

[9] Alonso-Coello P, Zhou Q, Martinez-Zapata MJ, Mills E, Heels-Ansdell D, Johanson JF, et al. Meta-analysis of flavonoids for the treatment of haemorrhoids. Br J Surg 2006;93:909−20.

[10] Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. BMJ 2006;333:597−600.

[11] Karanicolas PJ, Davies E, Kunz R, Briel M, Koka HP, Payne DM, et al. The pylorus: take it or leave it? Systematic review and meta-analysis of pylorus-preserving versus standard whipple pancreaticoduodenectomy for pancreatic or periampullary cancer. Ann Surg Oncol 2007;14:1825−34.

[12] Thavendiranathan P, Bagai A, Brookhart MA, Choudhry NK. Primary prevention of cardiovascular diseases with statin therapy: a meta-analysis of randomized controlled trials. Arch Intern Med 2006;166:2307−13.

[13] Armitage J. The safety of statins in clinical practice. Lancet 2007;370:1781−90.

ELSEVIER

Journal of Clinical Epidemiology 64 (2011) 401–406

Journal of Clinical Epidemiology

GRADE guidelines: 3. Rating the quality of evidence

Howard Balshem[a,*], Mark Helfand[a,b], Holger J. Schünemann[c], Andrew D. Oxman[d], Regina Kunz[e], Jan Brozek[c], Gunn E. Vist[d], Yngve Falck-Ytter[f], Joerg Meerpohl[g,h], Susan Norris[i], Gordon H. Guyatt[c]

[a]*Oregon Evidence-based Practice Center, Oregon Health and Science University, 3181 SW Sam Jackson Park Rd., Portland, OR 97239, USA*
[b]*Portland VA Medical Center, Portland, OR, USA*
[c]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[d]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St Olavs plass, 0130 Oslo, Norway*
[e]*Basel Institute of Clinical Epidemiology, University Hospital Basel, Hebelstrasse 10, 4031 Basel, Switzerland*
[f]*Division of Gastroenterology, Case Medical Center and VA, Case Western Reserve University, Cleveland, OH 44106, USA*
[g]*German Cochrane Center, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, 79104 Freiburg, Germany*
[h]*Division of Pediatric Hematology and Oncology, Department of Pediatric and Adolescent Medicine, University Medical Center Freiburg, 79106 Freiburg, Germany*
[i]*Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR 97239-3098, USA*

Accepted 30 July 2010

Abstract

This article introduces the approach of GRADE to rating quality of evidence. GRADE specifies four categories—high, moderate, low, and very low—that are applied to a body of evidence, not to individual studies. In the context of a systematic review, quality reflects our confidence that the estimates of the effect are correct. In the context of recommendations, quality reflects our confidence that the effect estimates are adequate to support a particular recommendation. Randomized trials begin as high-quality evidence, observational studies as low quality. ''Quality'' as used in GRADE means more than risk of bias and so may also be compromised by imprecision, inconsistency, indirectness of study results, and publication bias. In addition, several factors can increase our confidence in an estimate of effect. GRADE provides a systematic approach for considering and reporting each of these factors. GRADE separates the process of assessing quality of evidence from the process of making recommendations. Judgments about the strength of a recommendation depend on more than just the quality of evidence.

Keywords: Quality assessment; Body of evidence; Imprecision; Indirectness; Inconsistency; Publication bias

1. Introduction

In the two previous articles in this series, we introduced GRADE; provided an overview of the GRADE process for developing recommendations and the final outputs of that process, the evidence profile, and Summary of Findings table; and described the process for framing questions and identifying outcomes [1,2]. In this third article, we will introduce GRADE's approach to rating the quality of evidence. The goal is to provide a conceptual overview of the approach. A more detailed description, accompanied by examples, will follow in articles dealing with factors that may lead to rating down or rating up the quality of evidence [3–7].

2. What we do not mean by quality of evidence

In discussions of quality of evidence, confusion often arises between evidence and opinion and between quality of evidence and strength of recommendations. We, therefore, begin by explaining what we do not mean by quality of evidence.

3. Opinion is not evidence

In the absence of high-quality evidence, clinicians must look to lower quality evidence to guide their decisions.

The GRADE system has been developed by the GRADE Working Group. The named authors drafted and revised this article. A complete list of contributors to this series can be found on the *Journal of Clinical Epidemiology* Web site.

* Corresponding author. Oregon Evidence-based Practice Center, OHSU-BICC, 3181 SW Sam Jackson Park Rd., Portland, OR 97239, USA. Tel.: 503-220-8262 x54487; fax: 503-418-3332.
E-mail address: balshemh@ohsu.edu (H. Balshem).

0895-4356/$ - see front matter
doi: 10.1016/j.jclinepi.2010.07.015

Key Points

- GRADE provides a framework for assessing quality that encourages transparency and an explicit accounting of the judgments made.
- GRADE distinguishes between quality assessment conducted as part of a systematic review and that undertaken as part of guideline development.
- The optimal application of GRADE requires systematic review of the impact of alternative management strategies on all patient-important outcomes.
- Information about study limitations, imprecision, inconsistency, indirectness, and publication bias is necessary for decision makers, clinicians, and patients to understand and have confidence in the assessment of quality and estimate of effect size.

Confusion arises when, in such situations, guideline developers classify ''expert opinion'' as a type of evidence. Developing recommendations always requires the opinion of experts, the basis of which includes experience with patients, an understanding of biology and mechanism, and knowledge and understanding of preclinical and early clinical research as well as of the results of randomized clinical trials and observational studies. Guideline developers should always engage experts to help understand the evidence; they must also uncover and make clear the evidence that underlies the experts' opinions and rate the quality of that evidence, not the opinions that follow from the evidence and its interpretation.

An example illustrates the difference between evidence and expert opinion. Suppose that during attending rounds with medical students and residents, an endocrinologist explains the rationale for tight glycemic control in diabetes. Table 1 shows the two assertions he makes and the evidence he cites to support them. The evidence he cites for opinion 1 is exclusively his personal clinical experience. For opinion 2, he cites his own experience and refers (with no more than a general statement) to evidence from clinical research.

It seems highly plausible that opinion 1 might reasonably be based on careful observation. If patients who complain of fatigue, polyuria, or other symptoms return in a few days saying they are better, initiation of treatment is the likeliest explanation. The phenomenon of a patient who had no complaints returning, a few days later, to say how much better she is would be particularly memorable. Unfortunately, there are many other potential explanations of these observations. The endocrinologist's impression of the extent of patients' reports of benefit may be inaccurate, he may be forgetting many patients who failed to improve, or the apparent improvement in some patients may be because of natural history, placebo effects, leading questions on the part of the clinician, or the patient's desire to please. Without, at the very least, a rigorous and structured approach to data collection, we could consider the endocrinologist's report of his clinical experience (but not the opinion that he arrived at from his interpretation of that experience) as evidence from an uncontrolled case series and classify it as very low quality.

Whereas the implicit study design underlying the evidence for opinion 1 is a before–after study, opinion 2 suggests a parallel group comparison, which in this case has serious problems. If indeed his memory is accurate (patients with tighter control in his practice do achieve better outcomes), the reason may be that their success in controlling their glucose reflects differences in their underlying disease strongly associated with their likelihood of suffering complications. This risk of bias from unrecognized prognostic imbalance, as well as from the uncertainty and imprecision associated with the endocrinologist's memory of the events, would lead us again to classify his observations as very low quality evidence.

4. A particular quality of evidence does not necessarily imply a particular strength of recommendation

A second area of confusion relates to the distinction between assessing the quality of evidence and making a recommendation. Later articles in this series will provide a detailed discussion of GRADE's approach to deciding on the direction and strength of recommendations. We note here the importance of GRADE's explicit separation of the process for assessing the quality of a body of evidence from the process for making recommendations based in part on those assessments. Although higher quality evidence is more likely to be associated with strong recommendations than lower quality evidence, a particular level of quality does not imply a particular strength of recommendation. Sometimes, low or very low quality evidence can lead to a strong recommendation.

For instance, consider the decision to administer aspirin or acetaminophen to children with chicken pox. Observational studies have observed an association between aspirin administration and Reye's syndrome [8–11]. Because aspirin and acetaminophen are similar in their analgesic and antipyretic effects, the low-quality evidence regarding the potential harms of aspirin does not preclude a strong recommendation for acetaminophen.

Similarly, high-quality evidence does not necessarily imply strong recommendations. For example, faced with a first deep venous thrombosis (DVT) with no obvious provoking factor patients must, after the first months of anticoagulation, decide whether to continue taking warfarin long term. High-quality randomized controlled trials show that continuous warfarin will decrease the risk of recurrent thrombosis but at the cost of increased risk of bleeding and inconvenience [12–15]. Because patients with varying values and

Table 1
Expert opinion vs. evidence

Expert opinion	Evidence
Tight control will make a patient feel better	"In my 20 years in practice I have started treatment for newly diagnosed diabetes many times. I almost always see these patients back a week or so after starting treatment, and the great majority say they feel much better than they did before. Even a patient who denied having any complaints or symptoms will come back and say she has more energy, particularly in the afternoons, and will marvel at how much better she feels in general."
Tight control will reduce the long-term risk of developing kidney disease, neuropathy, and blindness	"I institute tight control on every patient—I believe they all deserve the best possible treatment—so I have a lot of experience with this. I have many patients who have been with me for a decade, or even several decades, and who take their medicine faithfully and have great blood sugars. These patients also have very few complications. On the other hand, I have a lot of patients who have terrible control and develop complications early on. Also, there are a lot of studies showing that tight control reduces the risk of complications."

preferences are likely to make different choices, guideline panels addressing whether patients should continue or terminate warfarin may—despite the high-quality evidence—offer a weak recommendation.

5. So what do we mean by "quality of evidence"?

GRADE distinguishes between quality assessment conducted as part of a systematic review and that undertaken in the process of guideline development. We, therefore, provide two definitions of "quality of evidence."

The optimal application of GRADE requires systematic reviews of the impact of alternative management approaches on all patient-important outcomes [1]. In the context of a systematic review, the ratings of the quality of evidence reflect the extent of our confidence that the estimates of the effect are correct. In the context of making recommendations, the quality ratings reflect the extent of our confidence that the estimates of an effect are adequate to support a particular decision or recommendation.

The reason for the different definitions is that the conduct of systematic reviews does not include processes required for making rigorous recommendations. In particular, unless the systematic review team includes members who will use the review as part of guideline development, authors of systematic reviews are, generally, not in a position to weigh the trade-offs between the desirable and undesirable consequences of adhering to a recommendation. Relevant stakeholders are in a better position to make these judgments. For example, in the DVT case described earlier, a systematic review might provide reliable estimates of the magnitude of effect and associated confidence intervals (CIs) for symptomatic thromboembolism and bleeding and the mortality associated with both of these events, but the reviewers who wrote it would not be able to provide reliable judgments about whether the benefit of warfarin treatment is worth the risk. Such judgments must also include considerations of values, cost, and pertinent stakeholder input.

On the other hand, a guideline (or a clinician applying the evidence from a systematic review) must assess the quality of the evidence in the context of the decision regarding anticoagulation. In considering this trade-off, a guideline panel must decide whether or not to recommend anticoagulation (and the strength of that recommendation) in light of the effect on the risk of symptomatic thromboembolism, their confidence in the effect estimates, and the corresponding risks and confidence in estimates of serious bleeding. Although the processes for assessing quality are the same, authors of systematic reviews and authors of guidelines will apply the criteria differently. We will highlight this different application of criteria in the fifth article in this series, which addresses the assessment of precision in rating the quality of the evidence [5].

6. Quality in GRADE means more than risk of bias

In the clinical epidemiological literature, when used at all, "quality" commonly refers to a judgment on the internal validity (i.e., risk of bias) of an individual study. To arrive at a rating, reviewers consider features in controlled trials such as randomization, allocation concealment, blinding, and use of intention to treat analysis. In observational studies, they consider appropriate measurement of exposure and outcome as well as appropriate control of confounding; in both controlled trials and observational studies, they consider loss to follow-up and may consider other aspects of design, conduct, and analysis that influence the risk of bias.

GRADE judgments refer not to individual studies but to a body of evidence, and quality, as used in GRADE, means more than risk of bias. A body of evidence (for instance, a number of well-designed and executed trials) may be associated with a low risk of bias, but our confidence in effect estimates may be compromised by a number of other factors (imprecision, inconsistency, indirectness, and publication bias). There are also factors, particularly relevant to observational studies, that may lead to rating up quality, including the magnitude of treatment effect and the presence of a dose–response gradient.

GRADE's specific uses of the terms "quality" and "risk of bias" (labeled "study limitations" in previous GRADE publications) require authors to take care in using these terms when they describe their findings and reasoning in

Table 2
Significance of the four levels of evidence

Quality level	Current definition	Previous definition
High	We are very confident that the true effect lies close to that of the estimate of the effect	Further research is very unlikely to change our confidence in the estimate of effect
Moderate	We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different	Further research is likely to have an important impact on our confidence in the estimate of effect and may change the estimate
Low	Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect	Further research is very likely to have an important impact on our confidence in the estimate of effect and is likely to change the estimate
Very low	We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect	Any estimate of effect is very uncertain

the context of a systematic review or guideline. Well-conducted studies may be part of a body of evidence rated low quality because they only provide indirect or imprecise evidence for the question of interest. Although clinical epidemiologists and others have attributed other meanings to the word "quality" (typically risk of bias), we believe the meaning described here corresponds more closely to the common and nontechnical understanding of "quality."

7. GRADE specifies four categories for the quality of a body of evidence

Although the quality of evidence represents a continuum, the GRADE approach results in an assessment of the quality of a body of evidence as high, moderate, low, or very low. Table 2 presents what GRADE means by each of these four categories and contrasts their current definition with the previous definition [16], which focused on the implications of the levels of evidence for future research (the lower the quality, the more likely further research would change our confidence in the estimates, and the estimates themselves). The earlier characterization has been criticized—we believe legitimately—because there are many situations in which we cannot expect higher quality evidence to be forthcoming. We, nevertheless, consider the prior characterization of quality to provide an alternative under circumstances when obtaining new compelling evidence is plausible.

8. Arriving at a quality rating

When we speak of evaluating quality, we are referring to an overall rating for each important outcome across studies. As discussed in the previous article in this series that addressed the framing of the question [2], before assessing the quality of the evidence, systematic reviewers and guideline developers should identify all potential patient-important outcomes, including benefits, harms, and costs. Reviewers will then assess the quality of evidence for each important outcome.

Table 3 summarizes GRADE's approach to rating the quality of evidence, which begins with the study design (trials or observational studies) and then addresses five reasons to possibly rate down the quality of evidence and three to possibly rate up the quality. Subsequent articles in this series will address, in detail, the meaning and use of each of these criteria. Here, we discuss why these criteria, in particular, have been identified as important in assessing the quality of a body of evidence.

Table 3
A summary of GRADE's approach to rating quality of evidence

Study design	Initial quality of a body of evidence	Lower if	Higher if	Quality of a body of evidence
Randomized trials	High ⇒	Risk of Bias −1 Serious −2 Very serious Inconsistency −1 Serious −2 Very serious Indirectness −1 Serious −2 Very serious Imprecision −1 Serious −2 Very serious Publication bias −1 Likely −2 Very likely	Large effect +1 Large +2 Very large Dose response +1 Evidence of a gradient All plausible residual confounding +1 Would reduce a demonstrated effect +1 Would suggest a spurious effect if no effect was observed	High (four plus: ⊕⊕⊕⊕) Moderate (three plus: ⊕⊕⊕◯) Low (two plus: ⊕⊕◯◯) Very low (one plus: ⊕◯◯◯)
Observational studies	Low ⇒			

9. Rationale for using GRADE's definition of quality

To be useful to decision makers, clinicians, and patients, systematic reviews must provide not only an estimate of effect for each outcome but also the information needed to judge whether these estimates are likely to be correct. What information about the studies in a review affects our confidence that the estimate of an effect is correct?

To answer this question, consider an example. Suppose you are told that a recent Cochrane review reported that, in patients with chronic pain, the number needed to treat (NNT) for clinical success with topical salicylates was 6 (95% CI = 4–13) compared with placebo. What additional information would you seek to help you decide whether to believe this estimate and how to apply it?

The most obvious questions might be the following: how many studies were pooled to get this estimate; how many patients did they include; and how wide were the CIs around the effect estimate? Were they randomized controlled trials? Did the studies have important limitations, such as lack of blinding or large or differential loss to follow-up in the compared groups? The questions thus far relate to GRADE categories of imprecision and risk of bias.

But there are also other important questions. Is there evidence that more studies of this treatment were conducted, but some were inaccessible to the reviewers? If so, how likely is it that the results of the review reflect the overall experience with this treatment? Did the trials have similar or widely varying results? Was the outcome measured at an appropriate time, or were the studies too short in duration to have much relevance? What part of the body was involved in the interventions (and thus, to what part of the body can we confidently apply these results)? These latter questions refer to the GRADE categories of publication bias, inconsistency, and indirectness. Without answers to (or at least information about) these questions, it is not possible to determine how much confidence to attach to the reported NNT and CIs.

GRADE identified its five categories—risk of bias, imprecision, inconsistency, indirectness, and publication bias—because they address nearly all issues that bear on the quality of evidence. For any given question, moreover, information about each of these categories is likely to be essential to judge whether the estimate is likely to be correct. These categories were arrived at through a case-based process by members of GRADE, who identified a broad range of issues and factors related to the assessment of the quality of studies. All potential factors were considered, and through an iterative process of discussion and review, concerns were scrutinized and solutions narrowed by consensus to these five categories.

GRADE's approach to quality implies that every systematic review should provide information about each of the categories (and any other pertinent issues in a particular case). Decision makers, whether they are guideline developers or clinicians, find it difficult to use a systematic review that does not provide this information. Good systematic reviews and clinical practice guidelines have commonly emphasized appraisal of the risk of bias (study limitations) using explicit criteria. Often, however, the focus has been on assessments across outcomes for each study rather than on each important outcome across studies. Assessment of other factors that determine how much confidence can be placed in estimates of effect has often been lacking. Before the adoption of GRADE, standards for reporting systematic reviews have not made clear how this information should be presented. GRADE provides a structure for systematic reviews and clinical practice guidelines to ensure they address the key questions that are pertinent to rating the quality of the evidence for all outcomes relevant to a particular question in a consistent systematic manner.

10. Conclusion

In closing, we caution against a mechanistic approach toward the application of the criteria for rating the quality of the evidence up or down. Although GRADE suggests the initial separate consideration of five categories of reasons for rating down the quality of evidence, and three categories for rating up, with a yes/no decision regarding rating up or down in each case, the final rating of overall evidence quality occurs in a continuum of confidence in the validity, precision, consistency, and applicability of the estimates. Fundamentally, the assessment of evidence quality is a subjective process, and GRADE should not be seen as obviating the need for or minimizing the importance of judgment or as suggesting that quality can be objectively determined.

As we repeatedly stress throughout this series, use of GRADE will not guarantee consistency in assessment, whether of the quality of evidence or of the strength of recommendations. There will be cases in which competent reviewers will have honest and legitimate disagreement about the interpretation of evidence. In such cases, the merit of GRADE is that it provides a framework that guides one through the critical components of this assessment and an approach to analysis and communication that encourages transparency and an explicit accounting of the judgments involved.

References

[1] Guyatt GH, Oxman AD, Kunz R, Vist GE, Brozek J, Norris S, et al. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383–94 [in this issue].

[2] Guyatt GH, Oxman AD, Kunz R, Atkins D, Brozek J, Vist GE, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395–400 [in this issue].

[3] Guyatt GH, Oxman AD, Vist GE, Kunz R, Brozek J, Alonso-Coello, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol 2011;64:407–15.

[4] Guyatt GH, Oxman AD, Montori V, Vist GE, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence - publication bias. J Clin Epidemiol. In press.

[5] Guyatt GH, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines: 6. Rating the quality of evidence—imprecision (random error). J Clin Epidemiol. In press.

[6] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence—inconsistency. J Clin Epidemiol. In press.

[7] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence—indirectness. J Clin Epidemiol. In press.

[8] Waldman RJ, Hall WN, McGee H, Van Amburg G. Aspirin as a risk factor in Reye's syndrome. JAMA 1982;247:3089–94.

[9] Starko KM, Ray CG, Dominguez LB, Stromberg WL, Woodall DF. Reye's syndrome and salicylate use. Pediatrics 1980;66:859–64.

[10] Halpin TJ, Holtzhauer FJ, Campbell RJ, Hall LJ, Correa-Villasenor A, Lanese R, et al. Reye's syndrome and medication use. JAMA 1982;248:687–91.

[11] Hurwitz ES, Barrett MJ, Bregman D, Gunn WJ, Pinsky P, Schonberger LB, et al. Public health service study of Reye's syndrome and medications: report of the main study. JAMA 1987;257:1905–11.

[12] Kearon C, Gent M, Hirsh J, Weitz J, Kovacs MJ, Anderson DR, et al. A comparison of three months of anticoagulation with extended anticoagulation for a first episode of idiopathic venous thromboembolism. N Engl J Med 1999;340:901.

[13] Campbell IA, Bentley DP, Prescott RJ, Routledge PA, Shetty HGM, Williamson IJ. Anticoagulation for three versus six months in patients with deep vein thrombosis or pulmonary embolism, or both: randomised trial. BMJ 2007;334:674.

[14] Kearon C, Ginsberg JS, Anderson DR, Kovacs MJ, Wells P, Julian JA, et al. Comparison of 1 month with 3 months of anticoagulation for a first episode of venous thromboembolism associated with a transient risk factor. J Thromb Haemost 2004;2:743–9.

[15] Agnelli G, Prandoni P, Santamaria MG, Bagatella P, Iorio A, Bazzan M, et al. Three months versus one year of oral anticoagulant therapy for idiopathic deep venous thrombosis. N Engl J Med 2001;345:165.

[16] Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 2008;336:924–6.

Sexology Today!

News and commentary from the fascinating science of sex, by Dr. James Cantor.

17 October 2018

American Academy of Pediatrics policy and trans- kids: Fact-checking

The American Academy of Pediatrics (AAP) recently published a policy statement entitled, *Ensuring comprehensive care and support for transgender and gender-diverse children and adolescents* (Rafferty, 2018). It was quite a remarkable document: Although almost all clinics and professional associations in the world use what's called the *watchful waiting* approach to helping GD children, the AAP statement rejected that consensus, endorsing only *gender affirmation.* With AAP taking such a dramatic departure from other professional associations, I was immediately curious about what evidence led them to that conclusion. (Extraordinary claims require extraordinary evidence, and all that.) As I read the works on which they based their policy however, I was pretty surprised…rather alarmed, actually: These documents simply did not say what AAP claimed they did. In fact, the references that AAP cited as the basis of their policy instead outright contradicted that policy, repeatedly endorsing *watchful waiting*.

The AAP statement was also remarkable in what it left out—namely, the outcomes research on GD children. There have been eleven follow-up studies of GD children, of which AAP cited one [Wallien & Cohen Kettenis (2008)], doing so without actually mentioning the outcome data it contained. The literature on outcomes was neither reviewed, summarized, nor subjected to meta-analysis to be considered in the aggregate—It was merely disappeared. (I have presented the complete list of the outcome studies on this blog before; they appear again at the bottom of this page together with their results, for reference.) As they make clear, *every* follow-up study of GD children, without exception, found the same thing: By puberty, the majority of GD children ceased to want to transition. AAP is, of course, free to establish whatever policy it likes on whatever basis it likes. But any assertion that their policy is based on evidence is demonstrably false, as detailed below.

AAP divided clinical approaches into three types—conversion therapy, watchful waiting, and gender affirmation. It rejected the first two and endorsed *gender affirmation* as the only acceptable alternative. Most readers will likely be familiar already with attempts to use conversion therapy to change sexual orientation. With regard to gender identity, AAP wrote:

> "[C]onversion" or "reparative" treatment models are used to prevent children and adolescents from identifying as transgender or to dissuade them from exhibiting gender-diverse expressions….Reparative approaches have been proven to be not only unsuccessful[38] but also deleterious and are considered outside the mainstream of traditional medical practice.[29, 39–42]

AAP's citations are:

38. Haldeman DC. The practice and ethics of sexual orientation conversion therapy. J Consult Clin Psychol. 1994;62(2):221–227
29. Adelson SL; American Academy of Child and Adolescent Psychiatry (AACAP) Committee on Quality Issues (CQI). Practice parameter on gay, lesbian, or bisexual sexual orientation, gender nonconformity, and gender discordance in children and adolescents. J Am Acad Child Adolesc Psychiatry. 2012;51 (9):957–974
39. Byne W. Regulations restrict practice of conversion therapy. LGBT Health. 2016;3(2):97–99
40. Cohen-Kettenis PT, Delemarrevan de Waal HA, Gooren LJ. The treatment of adolescent transsexuals: changing insights. J Sex Med. 2008;5(8):1892–1897

James M. Cantor, PhD

Dr. Cantor is a clinical psychologist and sexual behavior scientist. He has been studying and

41. Bryant K. Making gender identity disorder of childhood: historical lessons for contemporary debates. Sex Res Soc Policy. 2006;3(3):23–39
42. World Professional Association for Transgender Health. WPATH De-Psychopathologisation Statement. Minneapolis, MN: World Professional Association for Transgender Health; 2010. Available at: https://www.wpath.org/policies. Accessed April 16, 2017

These claims struck me as odd because *there are no studies of conversion therapy for gender identity.* Studies of conversion therapy have been limited to *sexual orientation*—specifically, the sexual orientation *of adults*—not *gender identity,* and not *children* in any case. The article AAP cited to support their claim (reference number 38) is indeed a classic and well-known review, but it is a review of *sexual orientation* research only. Neither gender identity, nor even children, received even a single mention in it. Indeed, the narrower scope of that article should be clear to anyone reading even just its title: "The practice and ethics of *sexual orientation* conversion therapy" (Haldeman, 1994, p. 221, italics added).

AAP continued, saying that conversion approaches for GD children have already been rejected by medical consensus, citing five sources. This claim struck me just as odd, however—I recalled associations banning conversion therapy for sexual orientation, but not for gender identity, exactly because there is no evidence for generalizing from adult sexual orientation to childhood gender identity. So, I started checking AAP's citations for that, and these sources too pertained only to sexual orientation, not gender identity (specifics below). What AAP's sources *did* repeatedly emphasize was that:

(1) Sexual orientation of adults is unaffected by conversion therapy and any other [known] intervention;
(2) Gender dysphoria in childhood before puberty desists in the majority of cases, becoming (cis-gendered) homosexuality in adulthood, again regardless of any [known] intervention; and
(3) Gender dysphoria in childhood persisting after puberty tends to persist entirely.

That is, in the context of GD children, it simply makes no sense to refer to externally induced "conversion": The majority of children "convert" to cisgender or "desist" from transgender *regardless* of any attempt to change them. "Conversion" only makes sense with regard to adult sexual orientation because (unlike childhood gender identity), adult homosexuality never or nearly never spontaneously changes to heterosexuality. Although gender identity and sexual orientation may often be analogous and discussed together with regard to social or political values and to civil rights, they are nonetheless distinct—with distinct origins, needs, and responses to medical and mental health care choices. Although AAP emphasized to the reader that "gender identity is not synonymous with 'sexual orientation'" (Rafferty, 2018, p. 3), they went ahead to treat them as such nonetheless.

To return to checking AAP's fidelity to its sources: Reference 29 was a practice guideline from the Committee on Quality Issues of the American Academy of Child and Adolescent Psychiatry (AACAP). Despite AAP applying this source to *gender identity,* AACAP was quite unambiguous regarding their intent to speak to *sexual orientation* and only to *sexual orientation:* "Principle 6. Clinicians should be aware that there is no evidence that *sexual orientation* can be altered through therapy, and that attempts to do so may be harmful. There is no established evidence that change in a predominant, enduring *homosexual* pattern of development is possible. Although sexual fantasies can, to some degree, be suppressed or repressed by those who are ashamed of or in conflict about them, sexual desire is not a choice. However, behavior, social role, and—to a degree—identity and self-acceptance are. Although operant conditioning modifies sexual fetishes, it does not alter *homosexuality*. Psychiatric efforts to alter *sexual orientation* through 'reparative therapy' *in adults* have found little or no change in *sexual orientation,* while causing significant risk of harm to self-esteem" (AACAP, 2012, p. 967, italics added).

Whereas AAP cites AACAP to support gender affirmation as the only alternative for treating GD children, AACAP's actual view was decidedly neutral, noting the lack of evidence: "Given the lack of empirical evidence from randomized, controlled trials of the efficacy of treatment aimed at eliminating gender discordance, the potential risks of treatment, and longitudinal evidence that gender discordance persists in only a small minority of untreated cases arising in childhood, further research is needed on predictors of persistence and desistence of childhood gender discordance as well as the long-term risks and benefits of intervention before any treatment to

teaching sexology, especially atypical sexualities, for over 20 years. His studies have been published in *Psychological Bulletin*, the *Journal of Abnormal Psychology*, and the *Journal of Consulting and Clinical Psychology*, and he served as Editor-in-Chief of *Sexual Abuse: A Journal of Research and Treatment*. He has appeared to discuss sexological issues on *CNN*, the *BBC*, *The New York Times*, and Dan Savage's *Savage Love*. He is Director of the Toronto Sexuality Centre and Associate Professor of Psychiatry at the University of Toronto. Summaries of his research and other projects are available at JamesCantor.org.

Top Posts

Do trans- kids stay trans- when they grow up?

Statistics faulty on how many trans- kids grow up to stay trans-?

On Russo's *Is there something unique about the transgender brain?* Well, yes and no.

How do I cure my pedophilia?

What's the difference between autogynephilia and trans women?

Receive new posts by email:

Email address... Submit

Currently, on twitter...

Tweets by @JamesCantorPhD

James Cantor Retweeted

Helen Joyce
@HJJoyceEcon

This is disgraceful. Imagine knowingly - or ignorantly- misrepresenting research on of all things Children's wellbeing. Weapons-grade professional malfeasance

2h

James Cantor Retweeted

Vote MidTerms Nov.6
@___lor__

Replying to @JamesCantorPhD @AmerAcadPeds

What's going on? I took two minutes to read it and the AAP policy contradicted the very materials it used as its reference. #ReadIt sexologytoday.org/2018/10/americ… @jrommel @Child_Neurology @TXProChild @ETRorg @Scott_Wiener @Prashaantpawar @KammMcKenzieOB @McMSchools @StressFreeKids

The American Academy of Pediatrics (AAP) recently published a policy statement entitled, *Ensuring comprehensive care and support for transgender and gender-diverse children and*

Embed View on Twitter

eliminate gender discordance can be endorsed" (AACAP, 2012, p. 969). Moreover, whereas AAP rejected watchful waiting, what AACAP recommended was: "In general, it is desirable to help adolescents who may be experiencing gender distress and dysphoria to defer sex reassignment until adulthood" (AACAP, 2012, p. 969). So, not only did AAP attribute to AACAP something AACAP never said, but also AAP withheld from readers AACAP's actual view.

Next, in reference 39, Byne (2016) also addressed only sexual orientation, doing so very clearly: "Reparative therapy is a subset of conversion therapies based on the premise that *same-sex attraction* are reparations for childhood trauma. Thus, practitioners of reparative therapy believe that exploring, isolating, and repairing these childhood emotional wounds will often result in reducing *same-sex attractions*" (Byne, 2016, p. 97). Byne does not say this of gender identity, as the AAP statement misrepresents.

In AAP reference 40, Cohen-Kettenis et al. (2008) did finally pertain to gender identity; however, this article never mentions conversion therapy. (!) Rather, in this study, the authors presented that clinic's lowering of their minimum age for cross-sex hormone treatment from age 18 to 16, which they did on the basis of a series of studies showing the high rates of success with this age group. Although it did strike me as odd that AAP picked as support against conversion therapy an article that did not mention conversion therapy, I could imagine AAP cited the article as an example of what the "mainstream of traditional medical practice" consists of (the logic being that conversion therapy falls outside what an 'ideal' clinic like this one provides). However, what this clinic provides is the very *watchful waiting* approach that AAP rejected. The approach espoused by Cohen-Kettenis (and the other clinics mentioned in the source—Gent, Boston, Oslo, and now formerly, Toronto) is to make puberty-halting interventions available at age 12 because: "[P]ubertal suppression may give adolescents, together with the attending health professional, more time to explore their gender identity, without the distress of the developing secondary sex characteristics. The precision of the diagnosis may thus be improved" (Cohen-Kettenis et al., 2008, p. 1894).

Reference 41 presented a very interesting history spanning the 1960s–1990s about how feminine boys and tomboyish girls came to be recognized as mostly pre-homosexual, and how that status came to be entered into the DSM at the same time as homosexuality was being *removed* from the DSM. Conversion therapy is never mentioned. Indeed, to the extent that Bryant mentions treatment at all, it is to say that treatment is entirely irrelevant to his analysis: "An important omission from the *DSM* is a discussion of the kinds of treatment that GIDC children should receive. (This omission is a general orientation of the DSM and not unique to GIDC)" (Bryant, 2006, p. 35). How this article supports AAP's claim is a mystery. Moreover, how AAP could cite a 2006 history discussing events of the 1990s and earlier to support a claim about the *current* consensus in this quickly evolving discussion remains all the more unfathomable.

Cited last in this section was a one-paragraph press release from the World Professional Association for Transgender Health. Written during the early stages of the American Psychiatric Association's (APA's) update of the DSM, the statement asserted simply that "The WPATH Board of Directors strongly urges the de-psychopathologisation of gender variance worldwide." Very reasonable debate can (and should) be had regarding whether gender dysphoria should be removed from the DSM as homosexuality was, and WPATH was well within its purview to assert that it should. Now that the DSM revision process is years completed however, history has seen that APA ultimately retained the diagnostic categories, rejecting WPATH's urging. This makes AAP's logic entirely backwards: That WPATH's request to depathologize gender dysphoria was *rejected* suggests that it is *WPATH's* view—and therefore, AAP policy—which fall "outside the mainstream of traditional medical practice." (!)

AAP based this entire line of reasoning on their belief that conversion therapy is being used "to prevent children and adolescents from identifying as transgender" (Rafferty, 2018, p. 4). That claim is left without citation or support. In contrast, what is said by AAP's sources is "delaying affirmation should *not* be construed as conversion therapy or an attempt to change gender identity" in the first place (Byne, 2016, p. 2). Nonetheless, AAP seems to appear to be doing exactly that: Simply relabeling non-gender affirmation models as conversion clinics.

Although AAP (and anyone else) may reject (what they label to be) conversion therapy purely on the basis of political or personal values, there is no evidence to back the AAP's stated claim about the existing science on gender identity at all, never mind gender identity of children.

AAP also rejected the watchful waiting approach, repeatedly calling it "outdated." The criticisms AAP provided, however, again defied the existing evidence, with even its own sources repeatedly calling that model the current standard. According to AAP:

> [G]ender affirmation is in contrast to the outdated approach in which a child's gender-diverse assertions are held as "possibly true" until an arbitrary age (often after pubertal onset) when they can be considered valid, an approach that authors of the literature have termed "watchful waiting." This outdated approach does not serve the child because critical support is withheld. Watchful waiting is based on binary notions of gender in which gender diversity and fluidity is pathologized; in watchful waiting, it is also assumed that notions of gender identity become fixed at a certain age. The approach is also influenced by a group of early studies with validity concerns, methodologic flaws, and limited follow-up on children who identified as TGD and, by adolescence, did not seek further treatment ("desisters").[45,47]

The citations from AAP's reference list are:

45. Ehrensaft D, Giammattei SV, Storck K, Tishelman AC, Keo-Meier C. Prepubertal social gender transitions: what we know; what we can learn—a view from a gender affirmative lens. Int J Transgend. 2018;19(2):251–268

47. Olson KR. Prepubescent transgender children: what we do and do not know. J Am Acad Child Adolesc Psychiatry. 2016;55(3):155–156.e3

I was surprised first by the AAP's claim that pubertal onset was somehow "arbitrary." The literature, including AAP's sources, repeatedly indicated the pivotal importance of puberty, noting that outcomes strongly diverge at puberty. According to AAP reference 29, in "*prepubertal* boys with gender discordance—including many without any mental health treatment—the cross gender wishes usually fade over time and do not persist into adulthood, with only 2.2% to 11.9% continuing to experience gender discordance" (Adelson & AACAP, 2012, p. 963, italics added), whereas "when gender variance with the desire to be the other sex is present *in adolescence,* this desire usually does persist through adulthood" (Adelson & AACAP, 2012, p. 964, italics added). Similarly, according to AAP reference 40, "Symptoms of GID *at prepubertal ages* decrease or even disappear in a considerable percentage of children (estimates range from 80–95%). Therefore, any intervention in childhood would seem premature and inappropriate. However, GID persisting *into early puberty* appears to be highly persistent" (Cohen-Kettenis et al., 2008, p. 1895, italics added). That follow-up studies of prepubertal transition differ from postpubertal transition is the very meaning of non-arbitrary. AAP gave readers exactly the reverse of what was contained its own sources. If AAP were correct in saying that puberty is an arbitrarily selected age, then AAP will be able to offer another point with as much empirical backing as puberty has.

Next, it was not clear on what basis AAP could say that watchful waiting withholds support—AAP cited no support for its claim. The people in such programs often receive substantial support during this period. Also unclear is on what basis AAP could already know exactly which treatments are "critical" and which are not—Answering that question is the very purpose of this entire endeavor. Indeed, the logic of AAP's claim appears entirely circular: If one were pre-convinced that the gender affirmation model is the only acceptable one, then watchful waiting withholds critical support only in the sense that it delays gender affirmation, the method one has already decided to be critical.

Although AAP's next claim did not have a citation appearing at the end of its sentence, binary notions of gender was mentioned both in references 45 and 47. Specifically, both pointed out that existing outcome studies have been about people transitioning from one sex to the other, rather than from one sex to an in-between or combination of masculine/feminine features. Neither reference presented this as a reason to reject the results from the existing studies of complete transition however (which is how AAP cast it). Although it is indeed true that the outcome data have been about complete transition, some future study showing that partial transition shows a different outcome for them would not invalidate what is known about complete transition. Indeed, data showing that partial transition gives better outcomes than

complete transition would, once again, support the watchful waiting approach which AAP rejected.

Next was a vague reference alleging concerns and criticisms about early studies. Had AAP indicated what those alleged concerns and flaws were (or which studies they were), then it would be possible to evaluate or address them. Nonetheless, the argument is a red herring: Because all of the later studies showed the same result as did the early studies, any such allegation is necessarily moot.

Reference 47 was a one-and-a-half page commentary which off-handedly mentions criticisms previously made of three of the eleven outcome studies of GD children, but does not provide any analysis or discussion (Olsen, 2016). The only specific claim was that studies (whether early or late) had limited follow-up periods—the logic being that had outcome researchers lengthened the follow-up period, then people who seemed to have desisted might have returned to the clinic as cases of "persistence-after-interruption." Although one could debate the merits of that prediction, AAP (and Olson) instead simply withheld from the reader the result from testing that prediction directly: Steensma and Cohen-Kettenis (2015) conducted another analysis of their cohort, by then ages 19–28 (mean age 25.9 years), and found that 3.3% (5 people of the sample of 150) later returned. That is, the childhood sample showing 70.0% desistence instead showed 66.7% desistance in long-term follow-up. It is up to the reader to decide whether that difference challenges the aforementioned conclusion that that majority of GD children cease to want to transition by puberty or represents a grasping at straws.

Reference

Steensma, T. D., & Cohen-Kettenis, P. T. (2015). More than two developmental pathways in children with gender dysphoria? *Journal of the American Academy of Child& Adolescent Psychiatry, 52,* 147–148.

Reference 45 did not support the claim that watchful-waiting is "outdated." Indeed, that source said the very opposite, referring to watchful waiting as the *current* approach: "Put another way, if clinicians are straying from SOC 7 guidelines for social transitions, not abiding by the watchful waiting model *favored by the standards,* we will have adolescents who have been consistently living in their affirmed gender since age 3, 4, or 5" (Ehrensaft et al., 2018, p. 255). Moreover, Ehrensaft et al. said there are cases in which they too would still use watchful waiting: "When a child's gender identity is unclear, the watchful waiting approach can give the child and their family time to develop a clearer understanding and is not necessarily in contrast to the needs of the child." Ehrensaft et al. are indeed critical of the watchful waiting model (which they feel is applied too conservatively), but they do not come close to the position the AAP policy espouses. Where Ehrensaft summarizes the potential benefits and potential risks both to transitioning and not transitioning, the AAP presents an ironically binary narrative.

In its policy statement, AAP told neither the truth nor the whole truth, committing sins both of commission and of omission, asserting claims easily falsified by anyone caring to do any fact-checking at all. AAP claimed, "This policy statement is focused specifically on children and youth that identify as TGD rather than the larger LGBTQ population" (p. 1); however, much of that evidence was about sexual orientation, not gender identity. AAP claimed, "Current available research and expert opinion from clinical and research leaders…will serve as the basis for recommendations" (p. 1-2); however, they provided recommendations entirely unsupported and even in direct opposition to that research and opinion.

AAP is advocating for something far in excess of mainstream practice and medical consensus. In the presence of compelling evidence, that would be exactly called for. The problems in Rafferty (2018), however, do not constitute merely a misquote, a misinterpretation of an ambiguous statement, or missing a reference or two. Rather, AAP's statement is a systematic exclusion and misrepresentation of entire literatures. Not only did AAP fail to provide *extraordinary* evidence, it failed to provide the evidence at all. Indeed, AAP's recommendations are *despite* the existing evidence.

Outcome Studies of GD Children and Their Results

Count Group	Study

2/16 4/16 10/16	gay trans-/crossdress straight/uncertain	Lebovitz, P. S. (1972). Feminine behavior in boys: Aspects of its outcome. *American Journal of Psychiatry, 128,* 1283–1289.
2/16 2/16 12/16	trans- uncertain gay	Zuger, B. (1978). Effeminate behavior present in boys from childhood: Ten additional years of follow-up. *Comprehensive Psychiatry, 19,* 363–369.
0/9 9/9	trans- gay	Money, J., & Russo, A. J. (1979). Homosexual outcome of discordant gender identity/role: Longitudinal follow-up. *Journal of Pediatric Psychology, 4,* 29–41.
2/45 10/45 33/45	trans-/crossdress uncertain gay	Zuger, B. (1984). Early effeminate behavior in boys: Outcome and significance for homosexuality. *Journal of Nervous and Mental Disease, 172,* 90–97.
1/10 2/10 3/10 4/10	trans- gay uncertain straight	Davenport, C. W. (1986). A follow-up study of 10 feminine boys. *Archives of Sexual Behavior, 15,* 511–517.
1/44 43/44	trans- cis-	Green, R. (1987). *The "sissy boy syndrome" and the development of homosexuality*. New Haven, CT: Yale University Press.
0/8 8/8	trans- cis-	Kosky, R. J. (1987). Gender-disordered children: Does inpatient treatment help? *Medical Journal of Australia, 146,* 565–569.
21/54 33/54	trans- cis-	Wallien, M. S. C., & Cohen-Kettenis, P. T. (2008). Psychosexual outcome of gender-dysphoric children. *Journal of the American Academy of Child and Adolescent Psychiatry, 47,* 1413–1423.
3/25 6/25 16/25	trans- lesbian/bi- straight	Drummond, K. D., Bradley, S. J., Badali-Peterson, M., & Zucker, K. J. (2008). A follow-up study of girls with gender identity disorder. *Developmental Psychology, 44*, 34–45.
17/139 122/139	trans- cis-	Singh, D. (2012). *A follow-up study of boys with gender identity disorder*. Unpublished doctoral dissertation, University of Toronto.
47/127 80/127	trans- cis-	Steensma, T. D., McGuire, J. K., Kreukels, B. P. C., Beekman, A. J., & Cohen-Kettenis, P. T. (2013). Factors associated with desistence and persistence of childhood gender dysphoria: A quantitative follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry, 52,* 582–590.

*For brevity, the list uses “gay” for “gay and cis-”, “straight” for “straight and cis-”, etc.

2 comments:

Hontas 18 October 2018 at 10:51

I am not a psychologist but as an academic citing sources that don't support what is being claimed should be reason for the retraction of a paper. I agree with what AAP is trying to do but this is not how to go about it. Academic integrity has to mean something. Could they be trying to reduce harm to those who DO turn out to be transgender as teens and adults, and their parents? In my humble opinion, "affirm" and give hormones probably isn't right either.

Hear me out.

Think of how the average, at least a little transphobic, parent is likely to implement "watchful waiting". Consider as an example the movie "Ma Vie En Rose". It could be said what they were doing in the beginning of the movie was a watchful waiting approach. "It's normal until about age 7," and all that. They were fine with Ludo dressing like a girl and having dolls and all of that until then.

The parents were told that if they watchfully waited the kid would likely "normally" stop. What happens when the kid does not stop? They loose it. They try to force them to stop being different. In real life that is often how trans gender youths end up as wards of the state. Part of watchful waiting has probably been a hope that the kid would not be trans in the end.

A month ago a well known transgender model posted a video about a frank discussion with her mother. When she had surgery her mother felt like her son "went on vacation then never came back". Her mother mourned. She still hoped somehow it wouldn't go that far.

That is the danger of watchful waiting for those who DO turn out to be transgender IF the people around them agreed in the hope they would desist.

Reply

Unknown 18 October 2018 at 12:21

@Hontas This is a mental health issue. Any medical interventions to treat mental health issues have huge risks, known and unknown, and are not "cures" for gender dysphoria. Medical action should never be positioned as the ultimate "end game" strategy for those who suffer--but that's not what you find in social media today, especially for children. If medical action is, after careful consideration, taken, some see a great reduction in their gender dsyphoria, some gain relief for a while before it returns in waves leading to more medical interventions b/c the last one was "not enough", and some get worse, some much worse. It depends. The motivations for the behavior are as varied as there are people, and the problem is that the AAP's take on "affirmation only" assumes only one action is correct. It's like saying all depressed people are depressed for the same reason and this is the one way to treat it. Basically, the AAP are guilty of being Distinction Deniers, and fail to see the harm they are exposing to others. How?

Well, for instance, the problem for us parents with teens with rapid onset gender dysphoria (ROGD) is that this is a social contagion highly influenced by social media. Teens and young adults have relatively easy--meaning no mental or medical health analysis required--access to cross sex hormones. Some can get them by mail. Some Planned Parenthoods have started giving out hormones with informed consent forms.

I'm sure there are some parents who are transphobic, but I'm more than sure that many parents do not want their child to take drastic medical interventions for something of dubious benefit when they want to give their child the gift of time to work things through and treat any underlying comorbid conditions.

Ironically, if your arguing that what AAP is doing is protecting that small number of children who will not desist after puberty (childhood GD is different from ROGD), you must also recognize that this will result in many false positives which encompasses ROGD teens and young adults. As the greatest predictor of persistence in an identity is social transition and that is fully endorsed in the affirmation model--and leads many of the afflicted to go on and seek medical "cures." Regrets will be many. These drugs are off-label use and do harm, visable and invisable. These kids are the experiment.

Reply

Enter your comment...

Comment as: Google Accour

Publish Preview

Links to this post

Create a Link

Home Older Post

Subscribe to: Post Comments (Atom)

App.0470

The opinions expressed here are my own and do not necessarily reflect those of any institution or organization with which I am affiliated.

Psychology of Sexual Orientation and Gender Diversity
2017, Vol. 4, No. 3, 341–353

2329-0382/17/$12.00 http://dx.doi.org/10.1037/sgd0000235

Suicide in Trans Populations: A Systematic Review of Prevalence and Correlates

Jay McNeil
Lancaster University

Sonja J. Ellis
Australian Catholic University

Fiona J. R. Eccles
Lancaster University

Trans people experience high rates of attempted suicide and suicidal ideation. No study to date has collated the various findings concerning correlates of trans suicide. This systematic review aimed to summarize the available data and provide recommendations based on this evidence. Articles were included if they were published before November 2016, were in English, were peer reviewed, and presented data concerning trans people's suicide attempts or ideation. Nine databases were searched, and 30 articles were selected. Discrimination emerged as strongly related to suicidal ideation and attempts, whereas positive social interactions and timely access to interventions appeared protective. Limitations included differences in how articles defined trans people or measured suicide and in their largely cross-sectional nature, making assumptions about causality in reference to lifetime ideation or attempts impossible. However, results clearly indicated a need to work at both individual and structural levels to reduce society- and service-level discrimination, enhance peer support, and ensure access to required interventions. The review highlights the need to explore suicidality in the trans population both in relation to general suicide models and in relation to models of minority stress.

Public Significance Statement
The findings presented here suggest that suicidality among trans people is complex, comprising a mix of individual, systemic, and structural factors. This article therefore highlights the importance of interpreting suicidal behavior in relation to specific models of minority stress and of working to address this issue across these different levels.

Keywords: transgender, suicidal ideation, suicide attempts, systematic review

The term *trans* is used to refer to a diverse range of people whose personal experience of gender is different from the conventional construction of gender as associated with the sex they were assigned at birth. Many people may be included under the umbrella term *trans,* such as trans men (those assigned female at birth but who identify as primarily masculine or male), trans women (those assigned male at birth but who identify as primarily feminine or female), those who define their gender as nonbinary (e.g., "bigender," "androgyne," "polygender"), and those who do not define their gender at all (e.g., "neutrois"). Trans people may have a fluid gender identity and may use more than one identity label at a time (e.g., *trans woman* and *genderqueer*) or a range of labels over time. Although some trans people may undergo medical interventions (e.g., hormone therapy, gender confirmation surgery) in relation to their gender identity, others may not. Similarly, some may socially transition (i.e., change their name and/or gender presentation).

Research and anecdotal evidence would seem to suggest that trans (or gender-diverse) people are at particular risk of suicide. For example, one United Kingdom study indicated that the lifetime prevalence of suicidal ideation may be as high as 84% (McNeil, Bailey, Ellis, Morton, & Regan, 2012), with 48% of ideators having attempted suicide (Bailey, Ellis, & McNeil, 2014). Similarly, del pozo de Bolger, Jones, Dunstan, and Lykins (2014) reported that within their sample 35% of trans people had attempted suicide in their lifetime. Rates have also been reported to be high in studies measuring suicidal ideation and attempts over the preceding year (e.g., Bauer, Pyne, Francino, & Hammond, 2013). Although there is a dearth of definitive data, rates of suicidal ideation and attempts in the trans population would appear to be substantially higher than the general population (e.g., 9.2%:

This article was published Online First July 24, 2017.

Jay McNeil, Division of Health Research, Lancaster University; Sonja J. Ellis, School of Psychology, Australian Catholic University; Fiona J. R. Eccles, Division of Health Research, Lancaster University.

This article is based on the doctoral thesis of Jay McNeil (now based at Lancaster, UK). The thesis was co-supervised by Sonja J. Ellis (at the time based at Sheffield Hallam University) and Fiona J. R. Eccles.

Correspondence concerning this article should be addressed to Fiona J. R. Eccles, Doctorate in Clinical Psychology, Furness College, Lancaster University, Lancaster LA1 4YG, United Kingdom. E-mail: jay@traverse-research.com

Nock et al., 2008; 5%: International Association for Suicide Prevention, 2012). These rates would also appear to be considerably higher than for other marginalized groups. For example, when compared to lesbian, gay, and bisexual (LGB) people, trans people were reported to be "162% more likely to have ever seriously considered committing suicide" (Irwin, Coleman, Fisher, & Marasco, 2014, p. 1181).

In the general population, the rate of suicidal ideation and attempts is related to the rate of completed suicide. In reviewing risk factors for suicide, Mościcki (2001) highlighted how having a history of suicide attempt is a substantial risk factor for later completed suicide. However, rates of completed suicide are difficult to ascertain for the trans population. It would be highly unusual for a person's trans status to be recorded on a death certificate; it is therefore not possible to ascertain trans suicide rates from coronial data. Conversely, because the trans population is hard to reach, published studies have tended to consist of limited, and typically self-selecting, samples, making it difficult to assess the extent of suicidality in the trans population as a whole. For example, the majority of studies have adopted a gender-binary approach to trans people, sampling only trans men and/or trans women, thus excluding nonbinary trans people. In other cases, sample selection is restricted to highly specific subgroups of the trans population, such as "trans gender women with a history of sex work" (e.g., Nemoto, Bödeker, & Iwamoto, 2011, p. 1) or solely those undergoing surgical intervention (e.g., Heylens, Verroken, De Cock, T'Sjoen & De Cuypere, 2014). Together these factors contribute to partial information and potential underreporting of suicide rates in trans populations (Bauer, Scheim, Pyne, Travers, & Hammond, 2015; Haas et al., 2010).

The negative impact that prejudice and discrimination have on the mental health and well-being (including increased risk of suicide) of individuals from marginalized groups has been well established (e.g., see Friedman, 1999; Meyer, 2003). Given the structural dominance of gender in society, trans people have a heightened consciousness of the extent to which their bodies and physical presentation do or do not conform to gender norms. Recent studies have shown that not only are trans people at significant risk of transphobic victimization but they have a heightened perception of this risk (Ellis, Bailey, & McNeil, 2016) and go to great lengths to avoid being victimized (Ellis, McNeil, & Bailey, 2014). In addition to social stigma, there are a number of personal challenges that trans people need to negotiate that also have a negative bearing on mental health and well-being. Factors such as distress in relation to gender, fears about transitioning, and delays or refusals in accessing gender confirmation interventions all contribute negatively to trans people's well-being, and these are viewed by many trans people as key factors in suicidality (e.g., see Bailey et al., 2014). As highlighted by Clements-Nolle, Marx, and Katz (2006), in LGB populations a clear link has been established between gender nonconformity and suicidality. Given that gender (as opposed to sexuality) is directly at stake for trans people, it is reasonable to assume that this would play a central part in placing trans people at risk of suicide.

Meyer's (2003) minority stress hypothesis (MSH) offers a more complex explanation for poor mental health in minority groups; it encompasses issues of social capital and microaggressions in its dividing of minority stressors into proximal and distal factors. In relation to suicide, minority populations experience the same risk factors as do majority group members (e.g., low income, not being married; Nock et al., 2008); however, they are also subject to additional stressors specific to their minority experience, which has an additive negative effect on mental health. Testa, Habarth, Peta, Balsam, and Bockting (2015) adapted the model for use with trans and gender-nonconforming people. They highlighted that trans-specific distal (external) factors would include gender-related discrimination, rejection and victimization, and nonaffirmation of someone's gender identity. Proximal (or internal) factors might include internalized transphobia, negative expectations, or concealment. The authors finally highlighted resilience or protective factors, including community connectedness and pride. Thus, the factors identified by Meyer (2003) have been easily adapted for trans populations. Although Meyer's minority stress hypothesis has recently been criticized for not considering the institutionalized nature of stressors affecting LGB and transgender (LGBT) individuals (e.g., see Riggs & Treharne, 2017) it still provides a useful framework for understanding suicidality at an individual level and is widely used in other studies.

Given the high rates of suicidal ideation and attempt among trans people, it is important to understand the factors contributing to this, and theories such as MSH offer a way of understanding how these factors may lead to such an outcome. Although research has explored trans people's suicidality, and a review that explores prevalence rates of suicidal ideation and attempt in these populations exists (E. Marshall, Claes, Bouman, Witcomb, & Arcelus, 2016), no systematic reviews have focused on collating the current evidence concerning correlates of these. Despite limitations in the breadth of samples (i.e., the bias toward binary trans participants), over the past 15 years there has been a proliferation of survey-based studies interfacing with trans suicide. Therefore, now is an appropriate time to undertake a cross-study examination of the determinants of suicide in trans populations. This review aims to summarize the evidence concerning factors that correlate with suicidal ideation and attempt in trans populations. The review also aims to provide recommendations for enhancing the future evidence base for supporting trans mental health.

Method

Inclusion and Exclusion Criteria

Studies included in this review consisted of those employing quantitative data about factors relating to suicidal ideation and/or attempts in trans people and that were published in English. Studies were excluded if they were qualitative or if they aggregated data about trans people with other populations (e.g., combining trans and LGB people's data).

Search Strategy

A literature search using a range of databases relevant to psychology and health sciences was conducted in November 2016. These databases included AHMED, Academic Search Complete, CINAHL, PsycInfo, PsycArticles, Web of Science, Scopus, OVID-EMBASE, and PubMed.

The following search terms, combined via the operator *OR,* were used to refer to trans people: *gender dysphoria, transgender, transsexual, gender variant, nonbinary, genderqueer, genderfluid,*

gender nonconformity, agender, two-spirit, kathoey, M2F, MTF, F2M, FTM, trans m, trans w*, male to female, female to male, androgyne, bigender, gender neutral, neutrois, bissu, kinnar, khusra, gender identity disorder.* The identity-related terms rather than diagnostic-related terms originated from different potential descriptions of trans people that were included in McNeil et al. (2012). In that study, gender-related options were determined by an advisory panel that included representation from different cultural trans communities, including researchers and clinicians with extensive theoretical and/or practical experience. Searches were conducted using free text terms rather than MeSH headings, because these headings did not return a sufficiently wide variety of the different terms that may be used to represent trans people.

The terms *attempted suicide, suicide prevention, suicide*,* and *suicidal ideation* were used to search specifically in relation to suicide. These terms were separated by the operator *OR*. The trans- and suicide-related search terms were then conjoined using the *AND* operator. Where it was possible, search terms were exploded and were unrestricted.

Selection

The process of screening the articles is presented in Figure 1. Initial searches returned 2,765 articles. Articles were screened by title, then abstract, leaving only 89 articles after duplicates were removed. Finally, full-text screening resulted in 30 articles, with no further articles matching all the criteria being identified from their references.

The articles identified during full-text screening were excluded if (a) they replicated data from the same sample as in another included study (e.g., Testa, Jimenez, & Rankin, 2014), (b) they were conference abstracts, or (c) the measures of suicidal ideation and attempt were unclear.

Quality Appraisal

The quality of reporting of the selected studies was established using criteria from the Strengthening the Reporting of Observa-

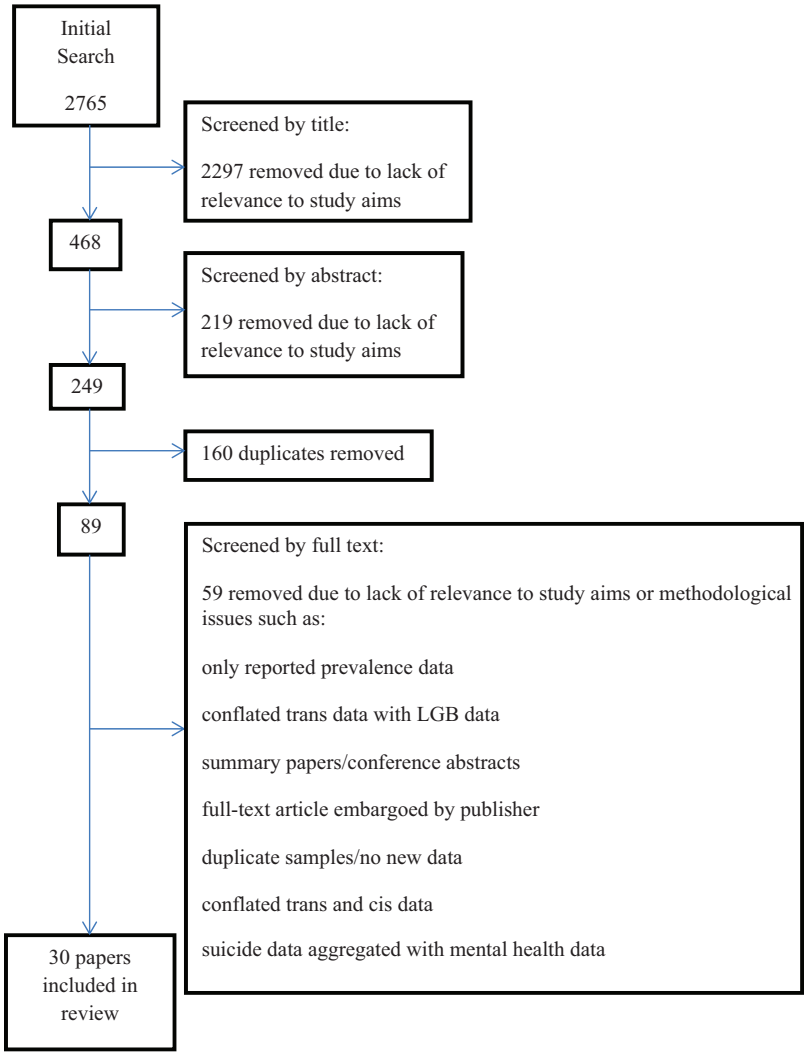


Figure 1. Flow diagram representing the number of articles removed at different stages of the review process. LGB = lesbian, gay, bisexual. See the online article for the color version of this figure.

tional Studies in Epidemiology (STROBE) checklist (Vandenbroucke et al., 2007), and the percentage of the relevant criteria met for each article was calculated. The quality of reporting among the articles varied considerably, from a low of 18% (Xavier, Bobbin, Singer, & Budd, 2005) to a high of 76% (Bauer et al., 2015). Most articles scored high for study design, methodology, and conclusions. Criteria that few met broadly related to how data were treated and analyzed and the provision of sufficient detail at each stage.

Results

Description of the Studies

Table 1 describes the main characteristics of the 30 studies reviewed here. The majority of studies ($n = 21$) were carried out in the United States, and among the remaining studies, two were carried out in Japan, two in South America (one in Argentina and one in Brazil), one in Canada, one in the United Kingdom, and the remaining three in countries of continental Europe (one in Belgium; one in Italy; and one in the Netherlands, Belgium, Germany, and Norway).

A key demographic for this study relates to gender. The articles varied in how they conceptualized their populations, with some representing only those who met diagnostic criteria for gender identity disorder (e.g., Heylens, Elaut, et al., 2014; Heylens, Verroken, et al., 2014; Lobato et al., 2007) and others allowing people to self-define their gender (e.g., Maguen & Shipherd, 2010; Nemoto et al., 2011). The definitions that participants could ascribe to were sometimes unclear; however, where stated, they differed between studies. For example, Goldblum et al. (2012) defined trans people as those

> having lived or wanting to live full-time in a gender opposite to their birth or physical sex; having or wanting to physically modify their body to match who they feel they really are inside; or having or wanting to wear the clothing of the opposite sex, in order to express an inner, cross-gender identity. (p. 470)

Alternatively, Nuttbrock et al. (2010) defined trans women as assigned "'male' at birth with a later conception of one's self as not 'completely male' in all situations or roles" (p. 14). For some articles, it was unclear how their population was defined or how participants were identified as trans (e.g., Colton Meier, Fitzgerald, Pardo, & Babcock, 2011). One article focused solely on trans men (Colton Meier et al., 2011), and four articles on trans women (Nemoto et al., 2011; Nuttbrock et al., 2010; Operario & Nemoto, 2005; E. C. Wilson, Chen, Arayasirikui, Wenzel, & Raymond, 2015). Sample sizes for individual studies varied from fewer than 100 (Grossman & D'Augelli, 2007) to over 4,000 (Miller & Grollman, 2015).

The reported demographic details varied substantially across studies. Age and some form of gender variable were provided by all articles. Given the importance and possible confounding nature of ethnicity or race and of sexual orientation, it is surprising that many articles omitted these variables. This, together with heterogeneity between articles concerning important variables (e.g., gender identity, country or culture of origin, how suicidal ideation and attempts were assessed), makes drawing conclusions across studies difficult. This level of variation needs to be borne in mind when considering the findings presented next.

Prevalence of Suicidal Ideation and Attempts

Prevalence rates among all studies varied; however, rates remained higher than for the general population, consistent with other studies. In the 17 articles that reported suicidal ideation, rates ranged from 37% (Mathy, 2003) to 83% (Testa et al., 2012). Rates of suicide attempt varied widely, ranging from 9.8% in a mixed trans group (Heylens, Verroken, et al., 2014) to 43% among trans men only (Colton Meier et al., 2011) and 44% in a mixed sample (Miller & Grollman, 2015). High prevalence rates are not entirely surprising for this population group, given the extent to which trans people routinely experience minority stress. Recent research (e.g., Grossman, Park, & Russell, 2016; Lehavot, Simpson, & Shipherd, 2016; Tebbe & Moradi, 2016) has indicated a strong association between minority stressors—for example, prejudice or discrimination, internalized anti-trans attitudes, fear of anti-trans stigma—and suicidality. As Tebbe and Moradi (2016) suggested, this highlights the importance of considering minority stressors (in addition to general factors) as risk factors for suicide in the trans population.

It is worth noting that the range reported in gender clinics varied from 9.8% (Heylens, Verroken, et al., 2014) to 21.2% (Colizzi, Costa, & Todarello, 2015), whereas for others it was 11.2% (Bauer et al., 2015) to 44% (Miller & Grollman, 2015). The highest rate from a gender clinic was recorded using a self-report questionnaire (Colizzi et al., 2015), as opposed to being gathered through interviews with clinicians directly determining care, which appeared to be the case for the other clinic articles. Suicide attempt may be underreported in gender clinics, because individuals are often aware that it may preclude access to medical transition interventions (Ellis, Bailey, & McNeil, 2015).

Correlates of Suicidal Ideation and Attempts

A number of demographic variables were associated with higher rates of suicide attempt, variables such as having a history of incarceration (Clements-Nolle et al., 2006) and lower socioeconomic status (e.g., Goldblum et al., 2012), whereas being in stable housing significantly decreased the odds of a lifetime suicide attempt (Lehavot et al., 2016; B. D. Marshall et al., 2016). Household income (Perez-Brumer, Hatsenbuehler, Oldenburg, & Bockting, 2015), relationship status (Maguen & Shipherd, 2010), sexual orientation (Clements-Nolle et al., 2006), and location (e.g., urban vs. rural area; Goldblum et al., 2012; B. D. Marshall et al., 2016; Perez-Brumer et al., 2015) were unrelated to suicide attempt. Variables unrelated to suicidal ideation included religion (Bauer et al., 2015) and relationship status (Terada et al., 2011). For other demographic variables, however, relationships appeared mixed or contradictory and are thus considered in more detail; these included gender, assigned sex, educational attainment, employment, age, and ethnicity.

The relationship between suicidal ideation and attempt and gender was complex. House, Van Horn, Coppeans, and Stepleman

Table 1
Reviewed Articles Exploring Factors Related to Suicide in Trans Populations

Article	Author(s)	Location	Aim	Sample size and composition
1	Bauer et al. (2015)	Canada	Identify intervenable social factors associated with a reduction in suicide risk	$N = 380$; trans women (47.4%), trans men (52.6%)
2	Clements-Nolle et al. (2006)	United States	Explore whether victimization and discrimination are independently associated with attempted suicide	$N = 515$; MTFs ($n = 392$) and FTMs ($n = 123$)
3	Colizzi et al. (2015)	Italy	Explore prevalence of dissociative disorders and symptoms before and after hormone therapy	$N = 118$; MTFs ($n = 82$), FTMs ($n = 36$)
4	Colton Meier et al. (2011)	Mainly United States	Provide evidence concerning the impact of testosterone on trans men's psychological state	$N = 369$; FTM transsexual-only sample
5	Effrig et al. (2011)	United States	Explore rates of harassment and discrimination, as well as mental health of college students, comparing those seeking treatment and those not	Sample 1 (not in or seeking counseling services): $n = 21,686$: trans people (gender unknown) and others ($n = 68$), cis women ($n = 13,244$), cis men ($n = 7,191$); Sample 2 (clinical sample seeking or receiving counseling services): $n = 27,616$: trans people (gender unknown; $n = 40$), cis women ($n = 16,615$), cis men ($n = 9,141$).
6	Goldblum et al. (2012)	United States	Explore the relation between gender-based victimization during school and attempted suicide	$N = 290$; trans women ($n = 147$; 33 with no plans to transition "full time"), trans men ($n = 81$; 29 with no transition plans)
7	Grossman and D'Augelli (2007)	United States	Explore whether "life-threatening behaviors" relate to parental reactions to participant's gender and feelings about their bodies	$N = 55$; young people: MTF ($n = 31$) and FTM ($n = 24$)
8	Grossman et al. (2016)	United States	Explore suicidal ideation and attempt in relation to perceived burdensomeness and thwarted belongingness	$N = 129$; young people: MTF ($n = 44$), FTM ($n = 40$), MTDG ($n = 14$), FTDG ($n = 31$)
9	Heylens, Elaut, et al. (2014)	The Netherlands, Belgium, Germany, Norway	Obtain rates of psychiatric diagnoses in people seeking gender reassignment who had also been diagnosed with gender identity disorder	$N = 305$; MTFs ($n = 182$) and FTMs ($n = 123$) diagnosed with gender identity disorder attending a gender clinic
10	Heylens, Verroken, et al. (2014)	Ghent, Belgium	Explore the psychological impact of different stages of medical gender reassignment interventions	$N = 57$; MTFs ($n = 46$) and FTMs ($n = 11$) undergoing gender confirmation surgery in a gender clinic
11	Hoshiai et al. (2010)	Japan	Investigate psychiatric comorbidity and life events in people attending the gender clinic	$N = 579$; MTF type ($n = 230$) and FTM type ($n = 349$) attending a gender clinic
12	House et al. (2011)	United States	Explore whether discrimination and interpersonal trauma relates to suicidal behaviors and self-injury in LGBT people	$N = 1,126$ LGBT people (164 trans people: 135 women or feminine people, 29 men or masculine people)
13	Kenagy and Bostwick (2005)	United States	Discuss trans community needs assessment findings	$N = 138$; MTFs ($n = 78$), FTMs ($n = 33$)
14	Lehavot et al. (2016)	United States	Explore correlates of suicidal ideation and suicide risk in transgender veterans	$N = 212$; trans women ($n = 186$), trans men ($n = 26$)
15	Lobato et al. (2007)	Brazil	Discuss the psychosocial characteristics of people using a clinic in Brazil	$N = 138$; MTFs ($n = 122$), FTMs ($n = 16$)
16	Maguen and Shipherd (2010)	United States	Report the frequency and predictive factors related to suicide in trans groups	$N = 153$; 125 people assigned male at birth (6% of whom had a male gender identity, 45% with a somewhat or entirely female identity), 28 people assigned female at birth (7% of whom had a female gender identity, 83% with a somewhat or entirely male identity)
17	B. D. Marshall et al. (2016)	Argentina	Examine the prevalence and correlates of suicide attempts in Argentinian trans people	$N = 482$; trans women ($n = 438$), trans men ($n = 44$)

(*table continues*)

Table 1 (*continued*)

Article	Author(s)	Location	Aim	Sample size and composition
18	Mathy (2003)	United States	Explore whether trans people are at increased risk of suicide, with LGB trans people being at an increased risk compared to heterosexual trans people; also explore whether suicidal trans people are more likely than nonsuicidal trans people to use support services and have compulsivity issues	N = 2,991; whole sample separated and reduced to comparison groups: heterosexual trans people (n = 29), nonheterosexual trans people (n = 44), heterosexual cis women (n = 1,083), lesbian cis women (n = 256), heterosexual cis men (n = 1,077), gay cis men (n = 356), psychosocially matched cis women (n = 73), psychosocially matched cis men (n = 73)
19	Miller and Grollman (2015)	United States	Explore whether perceived gender nonconformity is related to major and everyday transphobic discrimination and whether transphobic discrimination is related to attempted suicide; also explore whether transphobic discrimination mediates any relation between nonconformity and attempted suicide	N = 4,115; trans men (n = 1,601), trans women (n = 2,514).
20	Nemoto et al. (2011)	United States	Describe the impact of violence, transphobic events, and social support in trans women with a history of sex work, in relation to their ethnic or racial identity	N = 573; trans women with a history of sex work who identified as African American (n = 253), API (n = 110), Latina (n = 110), or White (n = 118)
21	Nuttbrock et al. (2010)	United States	Establish the impact of gender-related abuse across the lifetime, in particular its impact on depression and suicide	N = 571; MTF transgender people only
22	Operario and Nemoto (2005)	United States	Estimate HIV risk behaviors in API trans women	N = 110; API MTF transgender people
23	Perez-Brumer et al. (2015)	United States	Explore the relation between both individual and structural stigma and attempted suicide in a mixed trans group	N = 1,229; sample separated and coded as MTF (n = 697) or FTM (n = 532)
24	Rood et al. (2015)	United States	Explore whether transition, violence, and discrimination relate to suicidal ideation and whether transition and discrimination interact	N = 350; transgender women or MTF (n = 229), transgender men or FTM (n = 121)
25	Skagerberg et al. (2013)	United Kingdom	Describe suicidal behaviors and self-harm in young people prior to attending a gender clinic	N = 125; assigned natal male at birth (n = 68), assigned natal female at birth (n = 57)
26	Tebbe and Moradi (2016)	United States	Test the relation of minority stress and internalized transphobia with depression and suicide risk	N = 335; trans men (n = 90), trans women (n = 110), nonbinary (n = 128)
27	Terada et al. (2011)	Japan	Describe risk factors for suicidal ideation and self-harm in people attending a gender clinic	N = 500; MTF type (n = 189), FTM type (n = 311)
28	Testa et al. (2012)	United States	Explore whether physical violence and sexual assault relate to suicide and substance misuse	N = 271; trans women (n = 179), trans men (n = 92)
29	E. C. Wilson et al. (2015)	United States	Explore the impact that using different physical transition-related interventions has on the mental health of trans women	N = 314; trans women only
30	Xavier et al. (2005)	United States	Discuss trans community needs assessment findings	N = 248; assigned natal male at birth or MTFs (n = 188), and assigned natal female at birth or FTMs (n = 60)

Note. MTF = male to female; FTM = female to male; MTDG = male to different gender; FTDG = female to different gender; LGBT = lesbian, gay, bisexual, transgender; API = Asian/Pacific Islander. The terms used by the included studies to describe their samples have been reproduced here verbatim, and do not in all cases represent the terminology which the authors of this paper would endorse or use. The original terminology has been included to inform readers of the context in which trans people are considered in those studies.

(2011) reported that trans people and cis[1] women had higher rates of lifetime suicide attempt than did cis men and suggested that because most of their trans sample were women (82.3%), findings may reflect an effect of being female over being male. Similarly, Mathy (2003) found that suicidal ideation in trans people (although again greater than for cis people) was not different from that for psychosocially matched cis women or cis lesbians. It is unclear what the authors of that article intended in separating cis lesbians from cis women in general; however, it could be assumed that the intended demarcation was between cis lesbian women and cis

[1] cis refers to people whose gender identity is the same as the sex they were assigned at birth.

heterosexual women. Furthermore, Goldblum et al. (2012); Lehavot et al. (2016), and Maguen and Shipherd (2010) found that trans men were more likely to attempt suicide than were others, with Perez-Brumer et al. (2015) highlighting decreased odds for trans women compared with trans men. However, other studies found no differences between different gender identities in terms of ideation (Heylens, Elaut, et al., 2014; Hoshiai et al., 2010) or attempts (Clements-Nolle et al., 2006; Maguen & Shipherd, 2010; B. D. Marshall et al., 2016), and House et al. (2011) found trans women were more at risk of suicide attempt. There are important differences between the studies in terms of how participants were defined, where and how data were collected and gathered, the sizes of different trans populations included, and issues with comparing disaggregated cis people to mixed trans groups. Thus, although trans people consistently had higher rates of suicidal ideation and attempt than did cis people as a group, this relationship was not simple, and variations existed in subpopulations of each group.

Being assigned female rather than male at birth related to higher levels of suicidal ideation (Xavier et al., 2005) and suicide attempt (Goldblum et al., 2012; Maguen & Shipherd, 2010). However, in Skagerberg, Parkinson, and Carmichael's (2013) study, sex assigned at birth was not found not to be related to suicide attempt. However, these data were from clinic records rather than from participants, which may have affected the results in that participants may have purposefully minimized any difficulties out of concern for a potential impact on their access to treatment. Furthermore, their population was children and young people, who may not have made as many attempts as have adults. If being female assigned at birth relates to suicidal ideation and attempt, this may relate to the findings mentioned earlier, perhaps implying that others' perceptions of gender (regardless of an individual's identity) may lead to distressing experiences. However, there are differences in the assigned sex composition of the samples in these articles that make it difficult to draw firm conclusions.

Educational status appeared unrelated to suicidal ideation (Clements-Nolle et al., 2006) or suicide attempt (Clements-Nolle et al., 2006; B. D. Marshall et al., 2016); however, Perez-Brumer et al. (2015) did find that greater educational attainment related to decreased odds of lifetime suicide attempt. This may be an artifact of differences in the way educational attainment was coded, and further exploration would be useful.

Clements-Nolle et al. (2006) found that unemployment related to greater risk of suicide attempt; however, B. D. Marshall et al. (2016) and Terada et al. (2011) found that it was not related to suicide attempt or suicidal ideation, respectively. As such, the relationship between unemployment and suicide is unclear.

Age was another factor where there appeared to be considerable variation in findings. In terms of suicide attempt, B. D. Marshall et al. (2016) and Perez-Brumer et al. (2015) found no relationship with age, whereas Goldblum et al. (2012), Maguen and Shipherd (2010), and Nuttbrock et al. (2010) all found that younger age related to an increase in risk of suicide attempt. However, this relationship was not apparent for suicidal ideation. Terada et al. (2011) found that ideation was related to younger age among trans women, but age was unrelated in trans men. This may be at least partly explained by the results of Xavier et al. (2005), which demonstrated that suicidal ideation was high among those who were 13–19, then decreased substantially until around the age of 30, when it increased again. In Terada et al.'s sample, 58% of the trans women were over 30, compared to 26% of the trans men. Furthermore, only 26% of the trans women were under 24, whereas 42% of the trans men were under 24. Across all articles, the age profiles of participants differed, which may have contributed to the differences in findings, given that the relationship between age and suicidal ideation and attempt does not appear to be linear.

The relationship between suicidal ideation and ethnicity or race was relatively straightforward. Being White related to the highest levels of suicidal ideation (Kenagy & Bostwick, 2005; Nemoto et al., 2011), and African Americans were significantly less likely than others to disclose ideation (Xavier et al., 2005). For suicide attempt, the evidence was conflicted, with one study finding an increase related to being White (Clements-Nolle et al., 2006), whereas others found that multiracial and "other" groups, that is, those who were "non-White," reporting higher rates of suicide attempt than did White people (Goldblum et al., 2012; Perez-Brumer et al., 2015). Furthermore Maguen and Shipherd (2010) and Nemoto et al. (2011) found no relationship between ethnicity and suicide attempt, and the relationship highlighted in Perez-Brumer et al. (2015) for lifetime attempts was not present for attempts within the past 12 months. There may be differences between the samples that could affect these findings. For example, having a larger White cohort in one study (Goldblum et al., 2012) may have influenced the outcomes when compared to a study with a smaller White group and greater representation of people from other races (Clements-Nolle et al., 2006). However, one study that found no relationship between ethnicity and suicide attempt utilized a sample that was almost exclusively White (Maguen & Shipherd, 2010; 97% White).

Suicidality and Mental Health

Increased risks for ideation were seen in those who had a history of abuse (Grossman & D'Augelli, 2007) and in people with past or current use of either psychotherapy or medication (Mathy, 2003). Suicide attempt was related to having a history of drug or alcohol treatment (Clements-Nolle et al., 2006) and to psychiatric hospitalization even when demographic factors had been accounted for (Maguen & Shipherd, 2010), although a history of hospitalization predicted variance in attempts only once victimization-related variables were accounted for in the model. Higher rates of suicide attempt were evident in people with a diagnosis of dissociative identity disorder (Colizzi et al., 2015) and in those with low self-esteem (Clements-Nolle et al., 2006). No relationship was found between impulse-control difficulties and either suicidal ideation and attempt (Mathy, 2003) or having a diagnosed disorder (from Axis II of the *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.; American Psychiatric Association, 2000) and suicide risk (Heylens, Elaut, et al., 2014). Depression was related to increased suicidal ideation (Nemoto et al., 2011) and suicide attempt (Clements-Nolle et al., 2006). Help-seeking for distress was increased in those who were experiencing suicidal ideation; however, in terms of suicide attempt, there were no differences between those seeking help for distress and those not seeking treatment (Effrig, Bieschke, & Locke, 2011). House et al. (2011) found that psychiatric comorbidity was related to increased ideation in trans men but not in trans women; however, the sample

comprised a substantially larger group of trans men than trans women.

Trans-Related Variables and Suicidality

Many of the variables studied related to experiences that people had as a direct consequence of being trans. Suicide negativity (an index formed of questions relating to suicidal thoughts or feelings in relation to being LGBT; Grossman & D'Augelli, 2007) was associated with higher rates of suicidal ideation. Conformity with behaviors consistent with those expected based on the sex someone was assigned at birth (Grossman & D'Augelli, 2007) and age of onset of "gender dysphoria" (Heylens, Elaut, et al., 2014, p. 152) were not associated with suicidal ideation and attempt.

There was a complicated relationship between suicidal ideation and attempt and social support, which was mainly explored in Bauer et al. (2015). Having high levels of support was found to be strongly related to decreased suicidal ideation and attempt. When explored in more detail, there was no relationship between suicidal ideation and social support from people who were not the participants' parents; however, support from parents was related to decreased ideation. Unexpectedly, the authors found a positive relationship between higher levels of social support from leaders (e.g., employers or teachers) and increased suicide attempt, which they suggested may be due to attempts instigating increased support from those around the person, rather than causing it.

Bauer et al. (2015) demonstrated no relationship between undergoing some form of social transition and suicidal ideation. They did, however, find that having changed identity documents to match the gender someone identified as did relate to decreased suicidal ideation and attempt. When *transition* was defined as "living full-time in your gender of choice" (Rood, Puckett, Pantalone, & Bradford, 2015, p. 271), those planning to transition socially or who had already socially transitioned had elevated odds of reporting suicidal ideation compared to those who had no intention of transitioning. There was, however, a significant interaction between stage of transition and experiences of discrimination, whereby the greatest odds of reporting lifetime ideation resulted from planning to or having undergone transition and experiencing discrimination, whereas the lowest odds were among those who did not plan to transition and who did not experience discrimination.

Medical transition was similarly complex. Accessing transition-related medical care (or not) did not appear to relate to suicide attempt among trans women (E. C. Wilson et al., 2015). Heylens, Verroken, et al. (2014) also reported that in those who attended a gender clinic and underwent a medical transition, lifetime suicide attempt was not affected. However, this may be because lifetime rates are historical and cannot decrease. That they did not increase suggests that the interventions may have had an impact. However, one person did complete suicide during the pre- and postintervention measurements, which was not explored by the authors, because the person's data were excluded. It is important that findings from the gender clinics be considered with reference to their context as mediating access to medical interventions. In contrast, however, Bauer et al. (2015) reported less suicidal ideation for those undergoing a medical transition compared to those who were considering it. They further found that among those who were contemplating suicide, rates of suicide attempt increased during transition compared to other stages. There are differences in how *medical transition* was defined within these articles, which may account for some of this discrepancy; however, the samples were drawn from different populations (in terms of country location, as well as community vs. clinic), which may also limit comparison. Finally, being in a later stage of therapy (referring to an individual's medical process of transition) when presenting for an initial appointment at a gender clinic was related to increased suicidal ideation among trans women but not trans men (Terada et al., 2011).

In terms of the specific medical interventions that people might undergo, Colton Meier et al. (2011) found a small but nonsignificant decrease in lifetime prevalence of suicide attempt among a group of trans men who were taking hormones compared to those who were not. E. C. Wilson et al. (2015) found that hormones related to a significantly lower rate of suicidal ideation in trans women receiving them compared to those who were not. Overall, Bauer et al. (2015) reported that receiving hormones was associated with decreased suicidal ideation in a mixed group compared to those who had not started hormone therapy. E. C. Wilson et al. also studied breast augmentation among trans women and found that it related to lower levels of ideation in those who had undergone the procedure compared to those who had not, whereas whether genital surgery had been undertaken or not was unrelated to suicide attempt.

Negative Interpersonal Experiences

In general, experiencing lower levels of internalized transphobia related to lower levels of ideation (Bauer et al., 2015) and attempts (Bauer et al., 2015; B. D. Marshall et al., 2016; Perez-Brumer et al., 2015; although for Perez-Brumer et al., 2015, the relationship between internalized transphobia and suicide attempt became nonsignificant when referring to attempts within the past 12 months). Experiences of gender-related discrimination related to increased odds of suicidal ideation (Rood et al., 2015), although there was an interaction with transition (as discussed earlier) and suicide attempt (Clements-Nolle et al., 2006; Miller & Grollman, 2015). Lower levels of external transphobia related to lower levels of suicidal ideation and attempt (Bauer et al., 2015), whereas experiencing verbal victimization (Clements-Nolle et al., 2006) and gender-based victimization in school (Grossman & D'Augelli, 2007) all related to an increased risk of suicide attempt. For example, discrimination from health care staff and from the police related to increased odds of a lifetime attempt, although this did not hold when it was part of a multivariate model where internalized stigma remained significant (B. D. Marshall et al., 2016). Structural stigma (e.g., an environment with nonequal legislation for trans people) was also related to increased odds of lifetime attempts; however, it was not related to attempts within the past 12 months.

It is interesting that being seen by others as gender-nonconforming was significantly related to increased odds of lifetime suicide attempt, even when other variables were accounted for. However, this relationship was not present when transphobic discrimination was incorporated into the model. Thus, transphobic discrimination at least partly mediated the relationship between gender nonconformity and lifetime suicide attempt (Miller & Grollman, 2015).

Being victimized verbally (Clements-Nolle et al., 2006) or subjected to sexual violence (Clements-Nolle et al., 2006; Testa et al., 2012) or physical violence (Clements-Nolle et al., 2006; Maguen & Shipherd, 2010; Testa et al., 2012) also related to an increased risk of suicide attempt. However, Testa et al. (2012) found that the relationship between violence and identity may be more nuanced for suicidal ideation. Specifically, that for trans women physical violence related to increased ideation, whereas sexual violence did not; however, for trans men the converse was true, with sexual violence relating to increased ideation and there being no relationship with physical violence. Having experienced either physical or sexual violence related to suicidal ideation in Rood et al. (2015), with experiences of both associated with the highest odds of reporting ideation.

Discussion

The purpose of this article was to systematically review the published research on suicidality (i.e., suicidal ideation and attempt) in trans populations. The literature reviewed here consistently reported rates of suicide among trans people as being substantially higher than for the general population, although reported rates vary considerably between studies.

Across the literature reviewed, some demographic variables typically associated with suicidal ideation and attempts in the general population were not found to be associated with suicidality in trans people. For example, in the general population a relationship has been found between suicidal ideation and factors such as religious affiliation, relationship status, and lower levels of education (e.g., see Dervic et al., 2004; Lorant, Kunst, Huisman, Costa, & Mackenbach, 2005; Nock et al., 2008), yet in the studies reviewed there is limited evidence that these factors are related to suicidal ideation. There was also considerable inconsistency between the findings of different studies as to whether there was a relationship between suicidality and variables such as gender identity, sex assigned at birth, employment, age, and ethnicity.

The extent to which these findings appear contradictory suggest that a more nuanced exploration of the relationship between these factors and suicidality in the trans population is warranted. In particular, there is a lack of clarity around the extent to which particular risk factors may differentially impact on trans people as a function of key gender-related variables (e.g., sex assigned at birth, stage of transition). For example, some trans people whose gender identity and/or appearance is less congruent with established social norms may face ostracization and/or victimization due to others' intolerance of this, and these experiences may leave them at increased risk of suicide.

In relation to mental health, high rates of psychiatric diagnoses have been identified as associated with increased suicide risk in many populations (e.g., Qin, Agerbo, & Mortensen, 2003), including mixed LGBT groups (Irwin et al., 2014). Mental health related factors showed similar relationships here. For example, hospitalization is a significant suicide risk factor in the general population (Bostwick & Pankratz, 2000) and was also identified as a risk factor for suicide attempt in the trans population. In addition, a recent study (Tebbe & Moradi, 2016) has suggested that in the trans population depression is a mediating factor for suicide risk due to minority stress.

The complex and sometimes contradictory findings concerning demographic and mental health variables in these studies suggests that their relationships with suicidal ideation and attempt in trans communities are either different or more complex among trans people than for many other groups. This may be because of the additional pressures facing trans people conferred by virtue of their gender minority status.

In line with models of minority stress (Meyer, 2003; Riggs & Treharne, 2017), these findings highlight that discrimination and violence were consistently related to suicidal ideation and attempt in trans populations. Clements-Nolle et al. (2006) noted that many variables relating to suicide were similar to those with LGB people: "Societal risk factors such as [trans-related] gender-based discrimination and victimization are independently associated with attempted suicide" (p. 63). Furthermore, discrimination accounted for the relationship found between being perceived as gender-nonconforming and suicide attempt (Miller & Grollman, 2015) and was a factor in the relationship between suicidal ideation and transition status (Rood et al., 2015). The relationship between discrimination and negative outcomes was present at both the individual level (e.g., through interpersonal victimization) and the structural environmental level (e.g., through policy and legislation). However, as with other variables, gender differences did emerge, with physical and sexual trans-related violence having different impacts for trans men and trans women. Even so, gender differences in suicidality within the trans population have been underexplored, and where they were studied, explanations for these patterns have been unclear.

Testa et al., (2015) highlighted how distal variables such as discrimination may involve different stressors for trans people when compared to LGB people—for example, difficulties accessing legal recognition documents, accessing gender-appropriate medical care, or being unsafe in gendered spaces. They further posited an additional variable: "nonaffirmation" of identity. In relation to these factors, supporting people to live in a manner consistent with their gender, or affirming their identities legally through the provision of documentation or access to needed interventions, tended to relate to decreased suicidal ideation and attempt. However, it may be that people who were less distressed were more able to facilitate the changes they wished to make. Some aspects of physical transition were more complex; for example, although being in the process of transition was related to suicide attempts, this relationship involved other factors, such as discrimination. The variation in specific types of interventions that are either protective (e.g., hormone therapy; Bauer et al., 2015) or nonrelated (e.g., genital surgery; E. C. Wilson et al., 2015) may be a reflection of those that help the person interact in the world on a daily basis as the gender they are, compared to those that are more about internal consistency (thus the difference between the impact of hormones vs. genital surgery). Rood et al. (2015) demonstrated that undergoing or having undergone transition related to greater risks of lifetime ideation than for those not wanting to undergo transition. The authors suggested that their finding of an interaction between transition and discrimination may imply that transition confers a vulnerability to increased discrimination, thus explaining this interaction. Similarly, Bauer et al. (2015) suggested that although transition ultimately reduced risk, while undergoing it participants may face significant challenges that could add to their overall stress burden. This is commensurate with other re-

search (Bailey et al., 2014) showing that having undergone transition or being in the process of transition reflected a lower risk than did being prevented from transitioning.

Protective factors in the minority stress model largely relate to feeling connected to and part of a community, having social support, and having a sense of "identity pride." Here, identity development–related factors were unrelated to suicide (e.g., childhood gender conformity; Grossman & D'Augelli, 2007), although knowing people who were trans when first identifying as such seemed to be of some benefit, possibly through providing a positive concept of trans identities (Goldblum et al., 2012). Furthermore, having high levels of social support generally, and in particular parental social support, were protective in terms of suicide risk (Bauer et al., 2015). This finding echoes other research with trans populations, which found that rejection from family and peers increased the risk of suicide attempt (Haas, Rodgers, & Herman, 2014), and findings from studies with many other marginalized populations (e.g., Compton, Thompson, & Kaslow, 2005; Farrell, Bolland, & Cockerham, 2015).

Limitations and Considerations for Research and Policy

Distal stressors and social support have been studied in relation to suicidal ideation and attempt among trans groups, and the findings so far are consistent with those suggested by Meyer's (2003) minority stress theory. However, it would be useful to explore other minority stress factors, such as proximal factors (e.g., internalized transphobia, concealment) and protective factors (e.g., identity pride) to identify whether these variables can support a more cohesive understanding of suicide risk among trans people. That a substantial number of trans-specific and gender-based victimization variables have related to suicidal ideation and attempt (as protective and risk factors) demonstrates the added complexity of suicide risk in trans people over and above that experienced by the general population. Although models of minority stress (e.g., Meyer, 2003; Riggs & Treharne, 2017) offer valuable insights into additional factors impacting on trans suicide, the inclusion of other suicide models (e.g., interpersonal psychological theory or the clinical model; see Plöderl et al., 2014) may help in better understanding the impact of nonminority specific factors on suicidality in trans people and offer opportunities for more accurate comparison with general population samples.

One of the main barriers to understanding trans suicidality is the wide variation of samples and measures across the set of studies reviewed. For example, studies frequently used a mixture of validated (although it was not clear whether these had been validated for trans people) and researcher-designed measures; and data collection varied from interviewer-delivered surveys to computer-based delivery and clinical diagnostic assessments. Similarly, with some studies targeting samples from specific subgroups of the whole trans community and others aggregating data from across the trans population as a whole, there was a great deal of inconsistency in the samples. This level of variation adds too much complexity to the field, making it difficult to draw firm conclusions by cross-study comparison. Furthermore, the lack of clarity around correlates of suicidality, particularly gender, in the trans population suggests that a key consideration for future research is the avoidance of aggregating all trans people together. There has been ample research evidence that different subgroups have variable experiences of the world as a function of factors such as gender identity and stage of transition (e.g., see Ellis et al., 2014, 2016). It is also important to ensure that the experiences of trans people with nonbinary gender identities are appropriately included. Typically, nonbinary people have been either omitted from the research or subsumed into other categories. Given the central role that binary gender plays in the organization of the cis-centric society, specific attention to gender—and the experiences of those with a nonbinary gender identity in particular—could make a substantive contribution to the understanding of suicidality in the trans population as a whole.

As with the majority of work in the field of suicidality, the articles reviewed here focused on suicidal ideation and attempt rather than completed attempts. Gathering reliable data on completed suicides within the trans population would provide a more complete picture of trans suicide, perhaps shedding light on which subgroups of the trans population are at greatest risk. Currently data on completed trans suicides is not readily available, because a person's trans status is seldom recorded in coronial data.

In terms of the review methodology, only English-language articles were included, and within the studies there was a bias toward English-speaking participants and locations—and toward the United States in particular. These issues may have impacted upon the data; for example, Hjelmeland (2011) emphasized the need to take culture into account within suicide research. There are also differences in how trans people are viewed and treated (socially, legally, and medically) depending on location, which makes it difficult to draw comparisons. Ideally future research should involve liaison and joint working with trans organizations from a range of locations and countries to explore the cultural meanings of trans people's experiences.

Finally, an important consideration in light of the studies herein relates to cisnormativity (the tendency to assume that being cisgender is normative and that being transgender is therefore abnormal). Many articles referred to trans people by their assigned gender, which is disrespectful of the individuals the authors purported to be concerned with. Trans people were also compared to "men" and "women" rather than cis men or cis women, making their identities as men, women, or other nonbinary identities "other" to the "norm." Othering constitutes a microaggression, which ironically forms part of the complex interaction of factors contributing to poor mental health in trans people (e.g., Nadal, 2013). Therefore, greater care needs to be taken in future research both to refer to gender appropriately and, of importance, to specifically explore the part that othering and "misgendering" potentially play in mental well-being.

For future policy development, it is clear that lived experiences of discrimination and victimization have a substantial and devastating impact on trans people's lives. There are many areas where policy must be amended to improve inclusion and reduce discrimination of trans people. In particular, health policy needs reviewing, to enable trans people to more readily self-identify and access the support that they need (see Bailey et al., 2014; McNeil et al., 2012). This is important for commissioners, who also need to consider wider sources of research (e.g., research coming from the trans communities themselves) to avoid being exposed to only one dominant perspective that does not necessarily reflect the identities and realities of those for whom they may commission services.

Implications for Mental Health Practitioners

For psychologists, counselors, and others working with trans clients in the mental health space, these findings highlight the need to respond effectively and in a timely manner to suicidality. Rather than being an indicator of an underlying mental illness, suicidality in the trans population would appear to be attributable to unbearable stress resulting from a complex mix of risk factors (e.g., discrimination and victimization, social exclusion, identity concealment, internalized transphobia, decompensation) and a relative absence of protective factors (cf. Jobes, 2006). The focus of therapy therefore needs to be on working collaboratively with trans clients, focusing specifically on these issues and supporting them to develop resistance. However, given the substantial and potentially immediate risk of suicidality in this population, it is essential to gain an understanding of these underlying issues and assess suicide risk at the earliest opportunity. The collaborative assessment and management of suicidality (CAMS) model (Jobes, 2006) is a well-fitted approach to risk assessment and treatment planning in cases of this kind. An ecological approach to therapy (e.g., see E. R. Wilson, 2012) would also be useful for working with clients to build resilience and coping.

Although much can be accomplished in working with gender-diverse people at an individual level, it is important to also facilitate structural change. It may therefore be helpful to move beyond one-to-one work and therapeutic interventions that locate difficulties within the individual to also work creatively in fostering social support and in tackling the social causes of these disparities, which lie almost exclusively outside individual members of the trans community. Therapists have an ethical duty to challenge practice that may cause harm, such as delaying access to interventions to support gender affirmation or offering reparative or "conversion" therapies in relation to gender, and to ensure therapy does not inadvertently recreate the minority stressors experienced by trans clients in other settings (cf. Ellis et al., 2015).

Our recommendations are further echoed by professional practice guidance from the United States and the United Kingdom. Both the American Psychological Association's (2015) Guidelines for Psychological Practice With Transgender and Gender Nonconforming People and the British Psychological Society's (2012) Guidelines and Literature Review for Psychologists Working Therapeutically With Sexual and Gender Minority Clients place a similar emphasis on the importance of practicing in a trans-affirmative and culturally competent manner. In working with gender-diverse people, this would include recognition of and value for nonbinary identities, the importance of intersectionality, the value of reflection on one's own biases (and the potential impact of those biases), challenging social inequalities and enhancing trans-affirmative environments. Clearly, then, both the current evidence base and professional best practice point to the necessity of working at all levels, above that of one-to-one work directly with the individual, as an important strategy in improving individual client and community outcomes.

Conclusion

The findings presented here suggest that suicidality among trans people is incredibly complex, relating to multiple individual, systemic, and structural factors. It cannot be located solely within the individual, and any exploration of this must consider intersectional sources of oppression to fully capture its nuances.

To progress the understanding of suicidality in trans people, models such as Testa et al.'s (2015) adapted minority stress hypothesis and Riggs and Treharne's (2017) theory of decompensation offer a good starting point. However, a great deal more high-quality, thoughtful, and well-described research is required. In particular, greater attendance to group differences within the trans population is needed to help resolve some of the apparent contradictions between studies in the existing literature.

Ultimately the use of models around trans suicidality, and the consideration of factors that may or may not contribute to it, is useful practically only if it has a real-world impact in reducing distress and suicidality among gender-diverse people. It is clear at this stage that discrimination and victimization, both interpersonally and at a societal level, have a substantial role in suicidality. Steps must be taken to effect positive change, improve resilience factors such as access to and inclusion in trans organizations, and potentially save lives.

References

References marked with an asterisk are those that are included in the systematic review.

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders* (4th ed., text rev.). Washington, DC: Author.

American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist, 70,* 832–864. http://dx.doi.org/10.1037/a0039906

Bailey, L., Ellis, S. J., & McNeil, J. (2014). Suicide risk in the UK trans population and the role of gender transition in decreasing suicidal ideation and suicide attempt. *Mental Health Review, 19,* 209–220. http://dx.doi.org/10.1108/MHRJ-05-2014-0015

Bauer, G. R., Pyne, J., Francino, M. C., & Hammond, R. (2013). La suicidabilité parmi les personnes transe en Ontario: Implications en travail social et en justice sociale [Suicidality among trans people in Ontario: Implications for social work and social justice. *Service Social, 59,* 35–62. http://dx.doi.org/10.7202/1017478ar

*Bauer, G. R., Scheim, A. I., Pyne, J., Travers, R., & Hammond, R. (2015). Intervenable factors associated with suicide risk in transgender persons: A respondent driven sampling study in Ontario, Canada. *BMC Public Health, 15,* 525. http://dx.doi.org/10.1186/s12889-015-1867-2

Bostwick, J. M., & Pankratz, V. S. (2000). Affective disorders and suicide risk: A reexamination. *American Journal of Psychiatry, 157,* 1925–1932. http://dx.doi.org/10.1176/appi.ajp.157.12.1925

British Psychological Society. (2012). *Guidelines and literature review for psychologists working therapeutically with sexual and gender minority clients*. Retrieved February 23, 2016, from http://shop.bps.org.uk/guidelines-and-literature-review-for-psychologists-working-therapeutically-with-sexual-and-gender-minority-clients.html

*Clements-Nolle, K., Marx, R., & Katz, M. (2006). Attempted suicide among transgender persons: The influence of gender-based discrimination and victimization. *Journal of Homosexuality, 51,* 53–69. http://dx.doi.org/10.1300/J082v51n03_04

*Colizzi, M., Costa, R., & Todarello, O. (2015). Dissociative symptoms in individuals with gender dysphoria: Is the elevated prevalence real? *Psychiatry Research, 226,* 173–180. http://dx.doi.org/10.1016/j.psychres.2014.12.045

*Colton Meier, S. L., Fitzgerald, K. M., Pardo, S. T., & Babcock, J. (2011). The effects of hormonal gender affirmation treatments on mental health in female-to-male transsexuals. *Journal of Gay & Lesbian Mental Health, 15,* 281–299. http://dx.doi.org/10.1080/19359705.2011.581195

Compton, C. T., Thompson, N. J., & Kaslow, N. J. (2005). Social environment factors associated with suicide attempt among low-income African Americans: The protective role of family relationships and social support. *Social Psychiatry and Psychiatric Epidemiology, 40,* 175–185. http://dx.doi.org/10.1007/s00127-005-0865-6

del pozo de Bolger, A., Jones, T., Dunstan, D., & Lykins, A. (2014). Australian trans men: Development, sexuality, and mental health. *Australian Psychologist, 49,* 395–402. http://dx.doi.org/10.1111/ap.12094

Dervic, K., Oquendo, M. A., Grunebaum, M. F., Ellis, S., Burke, A. K., & Mann, J. J. (2004). Religious affiliation and suicide attempt. *American Journal of Psychiatry, 161,* 2303–2308. http://dx.doi.org/10.1176/appi.ajp.161.12.2303

*Effrig, J. C., Bieschke, K. J., & Locke, B. D. (2011). Examining victimization and psychological distress in transgender college students. *Journal of College Counseling, 14,* 143–157. http://dx.doi.org/10.1002/j.2161-1882.2011.tb00269.x

Ellis, S. J., Bailey, L. B., & McNeil, J. (2015). Trans people's experiences of mental health and gender identity services: A UK study. *Journal of Gay & Lesbian Mental Health, 19,* 4–20. http://dx.doi.org/10.1080/19359705.2014.960990

Ellis, S. J., Bailey, L., & McNeil, J. (2016). Transphobic victimisation and perceptions of future risk: A large-scale study of the experiences of trans people in the UK. *Psychology and Sexuality, 7,* 211–224. http://dx.doi.org/10.1080/19419899.2016.1181669

Ellis, S. J., McNeil, J., & Bailey, L. (2014). Gender, stage of transition and situational avoidance: A UK study of trans people's experiences. *Sexual and Relationship Therapy, 29,* 351–364.

Farrell, C. T., Bolland, J. M., & Cockerham, W. C. (2015). The role of social support and social context on the incidence of attempted suicide among adolescents living in extremely impoverished communities. *Journal of Adolescent Health, 56,* 59–65. http://dx.doi.org/10.1016/j.jadohealth.2014.08.015

Friedman, R. C. (1999). Homosexuality, psychopathology, and suicidality. *Archives of General Psychiatry, 56,* 887–888. http://dx.doi.org/10.1001/archpsyc.56.10.887

*Goldblum, P., Testa, R. J., Pflum, S., Hendricks, M. L., Bradford, J., & Bongar, B. (2012). The relationship between gender-based victimization and suicide attempts in transgender people. *Professional Psychology: Research and Practice, 43,* 468–475. http://dx.doi.org/10.1037/a0029605

*Grossman, A. H., & D'Augelli, A. R. (2007). Transgender youth and life-threatening behaviors. *Suicide and Life-Threatening Behavior, 37,* 527–537. http://dx.doi.org/10.1521/suli.2007.37.5.527

*Grossman, A. H., Park, J. Y., & Russell, S. T. (2016). Transgender youth and suicidal behaviors: Applying the interpersonal psychological theory of suicide. *Journal of Gay & Lesbian Mental Health, 20,* 329–349. http://dx.doi.org/10.1080/19359705.2016.1207581

Haas, A. P., Eliason, M., Mays, V. M., Mathy, R. M., Cochran, S. D., D'Augelli, A. R., . . . Clayton, P. J. (2010). Suicide and suicide risk in lesbian, gay, bisexual, and transgender populations: Review and recommendations. *Journal of Homosexuality, 58,* 10–51. http://dx.doi.org/10.1080/00918369.2011.534038

Haas, A. P., Rodgers, P. L., & Herman, J. L. (2014). *Suicide attempts among transgender and gender non-conforming adults: Findings of the National Transgender Discrimination Survey*. Retrieved July 1, 2015, from http://williamsinstitute.law.ucla.edu/wp-content/uploads/AFSP-Williams-Suicide-Report-Final.pdf

*Heylens, G., Elaut, E., Kreukels, B. P. C., Paap, M. C. S., Cerwenka, S., Richter-Appelt, H., . . . De Cuypere, G. (2014). Psychiatric characteristics in transsexual individuals: Multicentre study in four European countries. *British Journal of Psychiatry, 204,* 151–156. http://dx.doi.org/10.1192/bjp.bp.112.121954

*Heylens, G., Verroken, C., De Cock, S., T'Sjoen, G., & De Cuypere, G. (2014). Effects of different steps in gender reassignment therapy on psychopathology: A prospective study of persons with a gender identity disorder. *Journal of Sexual Medicine, 11,* 119–126. http://dx.doi.org/10.1111/jsm.12363

Hjelmeland, H. (2011). Cultural context is crucial in suicide research and prevention. *Crisis: The Journal of Crisis Intervention and Suicide Prevention, 32,* 61–64. http://dx.doi.org/10.1027/0227-5910/a000097

*Hoshiai, M., Matsumoto, Y., Sato, T., Ohnishi, M., Okabe, N., Kishimoto, Y., . . . Kuroda, S. (2010). Psychiatric comorbidity among patients with gender identity disorder. *Psychiatry and Clinical Neurosciences, 64,* 514–519. http://dx.doi.org/10.1111/j.1440-1819.2010.02118.x

*House, A. S., Van Horn, E., Coppeans, C., & Stepleman, L. M. (2011). Interpersonal trauma and discriminatory events as predictors of suicidal and nonsuicidal self-injury in gay, lesbian, bisexual, and transgender persons. *Traumatology, 17,* 75–85. http://dx.doi.org/10.1177/1534765610395621

International Association for Suicide Prevention. (2012). *World suicide prevention day: Facts and figures*. Retrieved from www.iasp.info/wspd/pdf/2012_wspd_facts_and_figures.pdf

Irwin, J. A., Coleman, J. D., Fisher, C. M., & Marasco, V. M. (2014). Correlates of suicide ideation among LGBT Nebraskans. *Journal of Homosexuality, 61,* 1172–1191. http://dx.doi.org/10.1080/00918369.2014.872521

Jobes, D. A. (2006). *Managing suicidal risk: A collaborative approach*. New York, NY: Guilford Press.

*Kenagy, G. P., & Bostwick, W. B. (2005). Health and social service needs of transgender people in Chicago. *International Journal of Transgenderism, 8,* 57–66. http://dx.doi.org/10.1300/J485v08n02_06

*Lehavot, K., Simpson, T. L., & Shipherd, J. C. (2016). Factors associated with suicidality among a national sample of transgender veterans. *Suicide and Life-Threatening Behavior, 46,* 507–524. http://dx.doi.org/10.1111/sltb.12233

*Lobato, M. I., Koff, W. J., Schestatsky, S. S., de Vasconcellos Chaves, C. P., Petry, A., Crestana, T., . . . Henriques, A. A. (2007). Clinical characteristics, psychiatric comorbidities and sociodemographic profile of transsexual patients from an outpatient clinic in Brazil. *International Journal of Transgenderism, 10,* 69–77. http://dx.doi.org/10.1080/15532730802175148

Lorant, V., Kunst, A. E., Huisman, M., Costa, G., & Mackenbach, J. (2005). Socio-economic inequalities in suicide: A European comparative study. *British Journal of Psychiatry, 187,* 49–54. http://dx.doi.org/10.1192/bjp.187.1.49

*Maguen, S., & Shipherd, J. (2010). Suicide risk among transgender individuals. *Psychology and Sexuality, 1,* 34–43. http://dx.doi.org/10.1080/19419891003634430

*Marshall, B. D., Socías, M. E., Kerr, T., Zalazar, V., Sued, O., & Arístegui, I. (2016). Prevalence and correlates of lifetime suicide attempts among transgender persons in Argentina. *Journal of Homosexuality, 63,* 955–967. http://dx.doi.org/10.1080/00918369.2015.1117898

Marshall, E., Claes, L., Bouman, W. P., Witcomb, G. L., & Arcelus, J. (2016). Non-suicidal self-injury and suicidality in trans people: A systematic review of the literature. *International Review of Psychiatry, 28,* 58–69. http://dx.doi.org/10.3109/09540261.2015.1073143

*Mathy, R. M. (2003). Transgender identity and suicidality in a nonclinical sample. *Journal of Psychology & Human Sexuality, 14,* 47–65. http://dx.doi.org/10.1300/J056v14n04_03

McNeil, J., Bailey, L., Ellis, S. J., Morton, J., & Regan, M. (2012). *Trans Mental Health Study 2012*. Retrieved from http://www.traverse-research.com/wp-content/uploads/2012/12/Trans-Mental-Health-2012.pdf

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: Conceptual issues and research evidence. *Psychological Bulletin, 129,* 674–697. http://dx.doi.org/10.1037/0033-2909.129.5.674

*Miller, L. R., & Grollman, E. A. (2015). The social costs of gender nonconformity for transgender adults: Implications for discrimination

and health. *Sociological Forum, 30,* 809–831. http://dx.doi.org/10.1111/socf.12193

Mościcki, E. K. (2001). Epidemiology of completed and attempted suicide: Toward a framework for prevention. *Clinical Neuroscience Research, 1,* 310–323. http://dx.doi.org/10.1016/S1566-2772(01)00032-9

Nadal, K. L. (2013). *That's so gay! Microaggressions and the lesbian, gay, bisexual, and transgender community*. http://dx.doi.org/10.1037/14093-000

*Nemoto, T., Bödeker, B., & Iwamoto, M. (2011). Social support, exposure to violence and transphobia, and correlates of depression among male-to-female transgender women with a history of sex work. *American Journal of Public Health, 101,* 1980–1988. http://dx.doi.org/10.2105/AJPH.2010.197285

Nock, M. K., Borges, G., Bromet, E. J., Alonso, J., Angermeyer, M., Beautrais, A., . . . Williams, D. (2008). Cross-national prevalence and risk factors for suicidal ideation, plans and attempts. *British Journal of Psychiatry, 192,* 98–105. http://dx.doi.org/10.1192/bjp.bp.107.040113

*Nuttbrock, L., Hwahng, S., Bockting, W., Rosenblum, A., Mason, M., Macri, M., & Becker, J. (2010). Psychiatric impact of gender-related abuse across the life course of male-to-female transgender persons. *Journal of Sex Research, 47,* 12–23. http://dx.doi.org/10.1080/00224490903062258

*Operario, D., & Nemoto, T. (2005). Sexual risk behavior and substance use among a sample of Asian Pacific Islander transgendered women. *AIDS Education and Prevention, 17,* 430–443. http://dx.doi.org/10.1521/aeap.2005.17.5.430

*Perez-Brumer, A., Hatzenbuehler, M. L., Oldenburg, C. E., & Bockting, W. (2015). Individual- and structural-level risk factors for suicide attempts among transgender adults. *Behavioral Medicine, 41,* 164–171. http://dx.doi.org/10.1080/08964289.2015.1028322

Plöderl, M., Sellmeier, M., Fartacek, C., Pichler, E.-M., Fartacek, R., & Kralovec, K. (2014). Explaining the suicide risk of sexual minority individuals by contrasting the minority stress model with suicide models. *Archives of Sexual Behavior, 43,* 1559–1570. http://dx.doi.org/10.1007/s10508-014-0268-4

Qin, P., Agerbo, E., & Mortensen, P. B. (2003). Suicide risk in relation to socioeconomic, demographic, psychiatric, and familial factors: A national register-based study of all suicides in Denmark, 1981–1997. *American Journal of Psychiatry, 160,* 765–772. http://dx.doi.org/10.1176/appi.ajp.160.4.765

Riggs, D. W., & Treharne, G. J. (2017). Decompensation: A novel approach to accounting for stress arising from the effects of ideology and social norms. *Journal of Homosexuality, 64,* 592–605. http://dx.doi.org/10.1080/00918369.2016.1194116

*Rood, B. A., Puckett, J. A., Pantalone, D. W., & Bradford, J. B. (2015). Predictors of suicidal ideation in a statewide sample of transgender individuals. *LGBT Health, 2,* 270–275. http://dx.doi.org/10.1089/lgbt.2013.0048

*Skagerberg, E., Parkinson, R., & Carmichael, P. (2013). Self-harming thoughts and behaviors in a group of children and adolescents with gender dysphoria. *International Journal of Transgenderism, 14,* 86–92. http://dx.doi.org/10.1080/15532739.2013.817321

*Tebbe, E. A., & Moradi, B. (2016). Suicide risk in trans populations: An application of minority stress theory. *Journal of Counseling Psychology, 63,* 520–533. http://dx.doi.org/10.1037/cou0000152

*Terada, S., Matsumoto, Y., Sato, T., Okabe, N., Kishimoto, Y., & Uchitomi, Y. (2011). Suicidal ideation among patients with gender identity disorder. *Psychiatry Research, 190,* 159–162. http://dx.doi.org/10.1016/j.psychres.2011.04.024

Testa, R., Habarth, J., Peta, J., Balsam, K., & Bockting, W. (2015). Development of the Gender Minority Stress and Resilience measure. *Psychology of Sexual Orientation and Gender Diversity, 2,* 65–77. http://dx.doi.org/10.1037/sgd0000081

Testa, R. J., Jimenez, C. L., & Rankin, S. (2014). Risk and resilience during transgender identity development: The effects of awareness and engagement with other transgender people on affect. *Journal of Gay & Lesbian Mental Health, 18,* 31–46. http://dx.doi.org/10.1080/19359705.2013.805177

*Testa, R. J., Sciacca, L. M., Wang, F., Hendricks, M. L., Goldblum, P., Bradford, J., & Bongar, B. (2012). Effects of violence on transgender people. *Professional Psychology: Research and Practice, 43,* 452–459. http://dx.doi.org/10.1037/a0029604

Vandenbroucke, J. P., von Elm, E., Altman, D. G., Gøtzsche, P. C., Mulrow, C. D., Pocock, S. J., . . . Egger, M. (2007). Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration. *Epidemiology, 18,* 805–835. http://dx.doi.org/10.1097/EDE.0b013e3181577511

*Wilson, E. C., Chen, Y.-H., Arayasirikul, S., Wenzel, C., & Raymond, H. F. (2015). Connecting the dots: Examining transgender women's utilization of transition-related medical care and associations with mental health, substance use, and HIV. *Journal of Urban Health, 92,* 182–192. http://dx.doi.org/10.1007/s11524-014-9921-4

Wilson, E. R. (2012). Assessment, diagnosis and treatment planning from the ecological perspective. In E. P. Cook (Ed.), *Understanding people in context: The ecological perspective in counselling* (pp. 179–206). Alexandria, VA: American Counseling Association.

*Xavier, J. M., Bobbin, M., Singer, B., & Budd, E. (2005). A needs assessment of transgendered people of color living in Washington, DC. *International Journal of Transgenderism, 8,* 31–47. http://dx.doi.org/10.1300/J485v08n02_04

Received February 7, 2017
Revision received June 1, 2017
Accepted June 3, 2017 ■

ELSEVIER

Journal of Clinical Epidemiology 66 (2013) 726−735

Journal of Clinical Epidemiology

GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength

Jeffrey C. Andrews[a,*], Holger J. Schünemann[b,c], Andrew D. Oxman[d], Kevin Pottie[e], Joerg J. Meerpohl[f,g], Pablo Alonso Coello[h,i], David Rind[j], Victor M. Montori[k], Juan Pablo Brito[k], Susan Norris[l], Mahmoud Elbarbary[m], Piet Post[n], Mona Nasser[o], Vijay Shukla[p], Roman Jaeschke[c], Jan Brozek[b], Ben Djulbegovic[q,r], Gordon Guyatt[b,c]

[a]*Vanderbilt Evidence-based Practice Center, Vanderbilt University, #27166-719 Thompson Lane, Nashville, TN 37204-3195, USA*
[b]*Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[c]*Department of Medicine, McMaster University, Hamilton, Ontario L8N 3Z5, Canada*
[d]*Norwegian Knowledge Centre for the Health Services, PO Box 7004, St. Olavs plass, Oslo 0130, Norway*
[e]*Department of Family Medicine, University of Ottawa, Ottawa, Canada*
[f]*German Cochrane Center, Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Berliner Allee 29, 79110 Freiburg, Germany*
[g]*Division of Pediatric Hematology and Oncology, Center for Pediatrics and Adolescent Medicine, University Medical Center Freiburg, Mathildenstrasse 1, 79106 Freiburg, Germany*
[h]*Iberoamerican Cochrane Center, CIBER de Epidemiología y Salud Pública, IIB, Sant Pau, Barcelona 08041, Spain*
[i]*Epidemiology and Public Health CIBER (CIBERESP), Hospital de la Sant Pau, Creu i Sant Pau, Barcelona 08041, Spain*
[j]*Harvard Medical School, Beth Israel Deaconess Medical Center, Healthcare Associates-E/Shapiro 6, 330 Brookline Aveen, Boston, MA 02215, USA*
[k]*Mayo Clinic, 200 1st SW St, Rochester, MN 55905, USA*
[l]*Department of Medical Informatics and Clinical Epidemiology, Oregon Health and Science University, Portland, OR 97239-3098, USA*
[m]*King Saud University for Health Sciences, Riyadh, Saudi Arabia*
[n]*Post Voor Zorg, Delft, The Netherlands*
[o]*Peninsula College of Medicine and Dentistry, Universities of Exeter and Plymouth, The John Bull Building, Tamar Science Park, Plymouth, PL68BU, UK*
[p]*Canadian Agency for Drugs and technologies in Health (CADTH), 600-865 Carling Avenue, Ottawa, Ontario K1S 5S8, Canada*
[q]*Division and Center for Evidence-Based Medicine and Health Outcomes Research, University of South Florida, Tampa, FL, USA*
[r]*H. Lee Moffitt Cancer Center & Research Institute, Tampa, FL, USA*

Accepted 3 February 2013; Published online 6 April 2013

Abstract

In the GRADE approach, the strength of a recommendation reflects the extent to which we can be confident that the composite desirable effects of a management strategy outweigh the composite undesirable effects.

This article addresses GRADE's approach to determining the direction and strength of a recommendation. The GRADE describes the balance of desirable and undesirable outcomes of interest among alternative management strategies depending on four domains, namely estimates of effect for desirable and undesirable outcomes of interest, confidence in the estimates of effect, estimates of values and preferences, and resource use. Ultimately, guideline panels must use judgment in integrating these factors to make a strong or weak recommendation for or against an intervention.

Keywords: GRADE; Quality of evidence; Strength of evidence; Guideline development; Recommendation; Evidence

1. Introduction

In prior articles in this series devoted to the GRADE approach to systematic reviews and practice guidelines, we have dealt with the process before developing recommendations, namely framing the question and choosing critical and important outcomes [1], rating the confidence in effect estimates for each outcome [2−8], dealing with resource

The GRADE system has been developed by the GRADE Working Group. The named authors drafted and revised this article. A complete list of contributors to this series can be found on the *Journal of Clinical Epidemiology* web site.

* Corresponding author. Tel.: (615) 343-5700.
E-mail address: jeff.andrews@vanderbilt.edu (J.C. Andrews).

0895-4356/$ - see front matter
http://dx.doi.org/10.1016/j.jclinepi.2013.02.003

use [9], rating the confidence in effect estimates across outcomes [10], and creating an evidence profile and a Summary of Findings table [11–13]. The immediately previous article described GRADE's approach to classifying the strength and direction of recommendations and discussed the implications of strong and weak recommendations, and the options for presentation and wording [14]. The present article presents GRADE's approach to moving from evidence to recommendations. As we did in the previous article, we will refer to guideline developers as "the panel."

1.1. Globalizing evidence and localizing decisions

The pithy summary by Eisenberg [15] on the relationship between evidence and recommendations, "globalize the evidence, localize the decisions," provides fundamental guidance for those working to produce evidence-based recommendations [15]. Summaries of evidence regarding alternative management strategies from the medical literature should ideally be very similar, no matter the site of the application of the recommendation.

Rating of confidence in estimates of effect (quality of evidence) may, however, differ for a variety of reasons. First, desirable and undesirable outcomes may be valued differently, leading to different thresholds of acceptability. This could lead to different judgments regarding imprecision, as we have highlighted in the article in this series dealing with imprecision [5].

Second, differences in values and preferences could lead to differences in the overall balance of desirable and undesirable outcomes and the rating of confidence in estimates: an outcome judged as critical by one panel (and thus included in the rating of overall confidence in estimates) may be judged important but not critical by another (and thus not included in the overall rating).

Finally, ratings of confidence may also differ as a result of uncertainties in the risk profile of untreated populations (baseline risk). We may be very confident of baseline risk in one setting but not at all confident in another. This could lead to rating down confidence in estimates for indirectness.

Continued rapid uptake of GRADE by organizations that produce systematic summaries of evidence will greatly facilitate the production of transparent evidence summaries. If organizations work together to produce summaries, there will be an enormous gain in efficiency [16]—even if, in the end, judgments about confidence in estimates will differ across settings, for reasons described in the preceding paragraphs.

We now turn to a systematic presentation of the determinants of direction and strength of recommendations.

2. Determinants of direction and strength of recommendations

GRADE has identified six determinants of the direction and strength of recommendations, namely the magnitude of estimates of effect of the interventions on important outcomes, confidence in those estimates, estimates of typical values and preferences, confidence in those estimates, variability of values and preferences, and resource use. In the presentation here, we will present these six determinants in four domains. We package magnitude of effect and typical values and preferences together with the label balance of desirable and undesirable consequences or "trade-offs." We also include uncertainty regarding typical values, and variability in values, in a single domain (Table 1).

Alternative groupings may work better, depending on the circumstances. We believe that the approach we present here is best for presenting the rationale for the recommendations to the guideline consumer audience. In developing recommendations, panels may want to keep all six determinants separate or group the three values and preferences determinants together.

Ultimately, guideline panels must integrate these six determinants to make a strong or weak recommendation for or against an intervention. Table 2 illustrates how the elements of the GRADE framework for moving from evidence to recommendations can be applied in making strong and weak recommendations, and Table 3 provides an example of the application in the management of chronic obstructive pulmonary disease.

2.1. Trade-offs between desirable and undesirable consequences of alternative management strategies

When we consider the balance between desirable and undesirable outcomes ("trade-offs"), we are considering two domains. The first is our best estimates of the magnitude of desirable effects and the undesirable effects. If a guideline panel has adhered to the GRADE process, they will find the best estimates of effect in the evidence profiles that they have prepared or accessed.

The second element that determines the balance among desirable and undesirable outcomes is the typical values that patients—or a population—apply to those outcomes. This can be otherwise conceptualized as the relative preferences for those outcomes—and thus the term we generally use, values and preferences (Box 1).

Ideally, to inform estimates of typical patient values and preferences, guideline panels will conduct or identify systematic reviews of relevant studies of patient values and preferences [18]. Given the paucity of empirical examinations of patients' values and preferences, however, well-resourced guideline panels will usually complement such studies with consultation with individual patients and patients' groups. The panel should discuss whose values these people represent, namely representative patients, a defined subset of patients, or representatives of the general population.

For example, the Canadian Collaboration for Immigrant and Refugees Health (CCIRH) guidelines sought to advance understanding of immigrant patient perspectives in

Table 1. Domains that contribute to the strength of a recommendation

Domains that contribute to the strength of a recommendation	Comment
Balance between desirable and undesirable outcomes (estimated effects), with consideration of values and preferences (estimated typical) (trade-offs)	The larger the differences between the desirable and undesirable consequences, the more likely a strong recommendation is warranted. The smaller the net benefit and the lower certainty for that benefit, the more likely a weak recommendation is warranted
Confidence in the magnitude of estimates of effect of the interventions on important outcomes (overall quality of evidence for outcomes)	The higher the quality of evidence, the more likely a strong recommendation is warranted
Confidence in values and preferences and variability	The greater the variability in values and preferences, or uncertainty in values and preferences, the more likely a weak recommendation is warranted
Resource use	The higher the costs of an intervention (the more resources consumed), the less likely a strong recommendation is warranted

two ways, namely they searched and synthesized evidence for immigrant perspectives in relation to each health condition, and worked closely with a community-based organization representing 18 ethnic groups to inform perceptions of immigrant patient perspectives [19]. Less well-resourced panels, without systematic reviews of values and preferences or consultation with patients and patient groups, must rely on unsystematic reviews of the available literature and their clinical experience of interactions with patients. How well such estimates correspond to true typical values and preferences is likely, in any particular situation, to be uncertain.

Whatever the source of estimates of typical values and preferences, explicit, transparent statements of the panel's choices are imperative. For example, in their recommendation regarding unmet contraceptive needs, the CCIRH attributed more value to supporting informed choice (empowerment) and less value to concern about causing couple and family discord [19]. Clinicians recognizing a family in which avoiding discord is paramount will therefore be aware that the recommendation is in that instance not appropriate.

Maximal explicitness requires quantification. For example, in the ninth iteration of the American College of Chest Physicians Antithrombotic Guidelines, the panel specified that they considered typical patients would value preventing one stroke equivalent to avoiding three serious gastrointestinal bleeds [18,20].

Having established their best estimates of typical values and preferences, a panel is in a position to assess the trade-off between the desirable and undesirable outcomes of an intervention vs. a comparator. The larger the gradient between the desirable and undesirable effects, the higher the likelihood that a panel will provide a strong recommendation. For example, the very large gradient between the benefits of low dose aspirin on reductions in death and

Table 2. Examples of strong and weak recommendation determinants

Factor	Example of strong recommendation	Example of weak recommendation
Balance between desirable and undesirable consequences of alternative management strategies. The closer the balance, the less likely a strong recommendation	Aspirin following myocardial infarction reduces mortality with minimal toxicity, inconvenience, and cost	Anticoagulation vs. aspirin in patients with atrial fibrillation with a $CHADS_2$ score of 1 (moderate risk of stroke); benefit in stroke reduction closely balanced with increased bleeding risk
Confidence in estimates of effect (quality of evidence). The lower the confidence, the less likely a strong recommendation	Many high quality randomized trials have shown the benefit of inhaled steroids in asthma	Only case series have examined the utility of pleurodesis in pneumothorax
Uncertainty or variability in values and preferences. The less the confidence in estimates of typical values and preferences, and the greater the variability, the less likely a strong recommendation	Relative confidence: evidence from empirical studies shows that patients place a substantially higher value on avoiding a debilitating stroke than on avoiding a serious gastrointestinal bleed	Uncertainty: there is no empirical evidence regarding the relative value patients place on avoiding a postoperative bleed that requires reoperation vs. a postoperative serious but nonfatal pulmonary embolus
	Little variability: young patients with lymphoma will invariably place a higher value on the life-prolonging effects of chemotherapy than on avoiding treatment toxicity	Greater variability: some older patients with lymphoma will place a higher value on the life-prolonging effects of chemotherapy than on avoiding treatment toxicity but others will not
Resource use. The higher the resource use, the less likely a strong recommendation	The low cost of aspirin vs. no antithrombotic prophylaxis against stroke in patients with transient ischemic attacks	The high cost of clopidogrel and of combination dipyridamole and aspirin vs. aspirin as prophylaxis against stroke in patients with transient ischemic attacks

Table 3. Evidence to recommendation framework: enhancing transparency when moving from evidence to recommendations

Question/recommendation: Should pulmonary rehabilitation vs. usual community care be used for COPD with recent exacerbation?			
Population: Patients with COPD and recent exacerbation of their disease			
Intervention: Pulmonary rehabilitation vs. no rehabilitation			
Setting (if relevant): Outpatient			
Decision domain	**Judgment**	**Reason for judgment**	**Subdomains influencing judgment**
Balance of desirable and undesirable outcomes Given the best estimate of typical values and preferences, are you confident that the benefits outweigh the harms and burden or vice versa?	Yes ☒ No ☐	The desirable consequences are substantial (including substantial reduction in hospitalization, small but important reduction in mortality, and improvement in quality of life that exceeds the minimal important difference) and valued highly. The undesirable consequences, inconvenience, and burden are relatively minor and associated with minimal disutility.	Baseline risk for desirable and undesirable outcomes: • Is the baseline risk similar across subgroups? • Should there be separate recommendations for subgroups? Relative risk for benefits and harms: • Are the relative benefits large? • Are the relative harms large? Requirement for modeling: • Is there a lot of extrapolation and modeling required for these outcomes? Typical values: • What are the typical values? • Are there differences in the relative value of the critical outcomes?
Confidence in estimates of effect (quality of evidence) Is there high or moderate quality evidence?	Yes ☒ No ☐	⊕⊕⊕O There is moderate-(mortality, function, and quality-of-life outcomes)-to-high (hospitalizations) quality evidence for the desirable consequences, and quality evidence for the undesirable (burden)	Confidence in estimates of benefits and downsides, confidence in estimates of resource use. Consider all critical outcomes, including the possibility that some may not be measured. Key reasons for rating evidence down or rating up
Values and preferences Are you confident about the typical values and preferences and are they similar across the target population?	Yes ☒ No ☐	We can be confident that patients place a high value on avoiding hospitalizations and mortality as well as improving quality of life and a low value on avoiding the inconvenience associated with rehabilitation. We can be confident that these values vary little among patients with chronic respiratory disease.	Source of typical values (panel or study of general population or patients) Source of estimates of variability and extent of variability Method for determining values satisfactory for this recommendation
Resource implications Are the resources worth the expected net benefit from following the recommendation?	Yes ☒ No ☐	There are resources required to provide pulmonary rehabilitation but these are balanced by decreased resource needs as a result of decreased hospitalizations and net cost is well worth it given the desirable outcomes.	What are the costs per resource unit? Feasibility: • Is this intervention generally available? Opportunity cost: • Is this intervention and its effects worth withdrawing or not allocating resources from other interventions Differences across settings: • Is there lots of variability in resource requirements across settings?
Overall strength of recommendation	Strong	The guideline panel recommends that patients with recent exacerbations of their COPD undergo pulmonary rehabilitation (Note: this is a hypothetical recommendation developed for this article and not intended for clinical decision making).	
Evidence to recommendation synthesis	The moderate-to-high confidence in the moderate-to-large magnitude of effects on highly valued outcomes, and the moderate-to-high confidence that undesirable outcomes are modest and their avoidance not highly valued suggest a strong recommendation.		

Abbreviation: COPD, chronic obstructive pulmonary disease.

Box 1 Terminology for "values and preferences"

Values and preferences is an overarching term that includes patients' perspectives, beliefs, expectations, and goals for health and life [17]. More precisely, they refer to the processes that individuals use in considering the potential benefits, harms, costs, limitations, and inconvenience of the management options in relation to one another. For some, the term "values" has the closest connotation to these processes. For others, the connotation of "preferences" best captures the notion of choice. Thus, we use both words together to convey the concept.

recurrent myocardial infarction (MI) after an MI [21] and the undesirable consequences of minimal side effects and costs make a strong recommendation very likely (Table 2).

In contrast, the narrower the magnitude of the gradient between desirable and undesirable consequences, the higher the likelihood that a guideline panel will make a weak recommendation. For instance, consider the choice of immunomodulating agents, namely cyclosporine and tacrolimus in kidney transplant recipients [22]. Tacrolimus results in better graft survival (a highly valued outcome), but at the important cost of a higher incidence of diabetes (the long-term complications of which can be devastating).

Table 2 presents a second example of a close trade-off in which patients with atrial fibrillation typically are more stroke averse than bleeding averse. If, however, the risk of stroke is sufficiently low, the trade-off between stroke reduction and increase in bleeding risk with anticoagulants is closely balanced.

Without considering the associated values and preferences, assessing large vs. small magnitude of effects may be misleading. For instance, in patients with cancer, chemotherapeutic agents may have large (albeit temporary) adverse effects such as nausea, fatigue, hair loss, and paresthesias. The chemotherapy may have only a small effect on reducing mortality. Despite the discrepancy in magnitude of effect, most patients may choose chemotherapy because of the very high value they place on a small mortality reduction.

2.2. Uncertainty and variability in values and preferences

We have noted that systematic study of patients' values and preferences are very limited. As a result, panels will often be uncertain about typical values and preferences. The greater is that uncertainty, the more likely they will make a weak recommendation.

Given the sparse systematic study of patients' values and preferences, one could argue that large uncertainty always exists about the patients' perspective. On the other hand, some systematic study of values and preferences and decision making has been completed, and clinicians' experience with patients may provide considerable additional insight.

Indeed, on occasion, panels will, on the basis of clinical experience, be confident regarding typical patient's values and preferences. Pregnant women's strong aversion to even a small risk of important fetal abnormalities may be one such situation [20].

A second concern that may make a weak recommendation more likely is large variability in values and preferences. To the extent large variability exists, it is less likely that a single recommendation would apply uniformly across all patients, and the right course of action is likely to differ between patients.

Empirical evidence may inform estimates of variability in recommendations. For instance, Devereaux et al. [23] asked patients at risk of atrial fibrillation how many serious gastrointestinal bleeds they would tolerate and still be willing to use an anticoagulant to prevent a stroke. Although most patients placed a high value on avoiding a stroke and were ready to accept a bleeding risk of 22% to reduce their chances of having a stroke by 8%, diversity in values and preferences was also apparent. A few patients were ready to accept only a small risk of bleeding to reduce their stroke risk by 8%. These data, consistent with other studies of values and preferences regarding anticoagulation in atrial fibrillation [18], suggest that only in patients at appreciable risk of stroke would a strong recommendation for warfarin be warranted.

Although systematic study will lead to the highest confidence, panelists may express confidence in their estimates of variability in values and preference on the basis of clinical experience. In the example cited earlier, clinicians may be confident not only that the typical expectant mother will have a strong aversion to even a small risk of important fetal abnormalities but also that these values and preferences are virtually uniform across the population.

On the other hand, clinical experience may leave a panel confident that values and preferences differ widely among patients. For example, clinical experience makes it clear that an expectant couples' desire to undergo a genetic test that increases the risk of spontaneous miscarriage will differ greatly depending on their willingness to act on knowledge about a fetal anomaly and their attitude toward the loss of a normal pregnancy. Situations such as these when recommendations are particularly dependent on differing values and preferences may dictate, in addition to making a weak recommendation, including descriptions of how varying values and preferences will determine the optimal decision [14].

A hopeful patient may place more emphasis on a small chance of benefit, whereas a pessimistic, risk-averse patient may place more emphasis on avoiding the risks associated with a potentially beneficial therapy. Some patients may

have a belief that even if the risk of an adverse event is low, they will be the person who will suffer such an adverse effect.

For example, in patients with idiopathic pulmonary fibrosis, evidence for the benefit of steroids warrants only low confidence, whereas we can be very confident of a wide range of adverse effects associated with steroids. The hopeful patient with pulmonary fibrosis may be enthusiastic about use of steroids, whereas the risk-averse patient is likely to decline.

2.3. Confidence in estimates of effect (quality of evidence)

Another determinant of the direction and strength of recommendations is our confidence in the estimates of effect. Typically, a strong recommendation is associated with high, or at least moderate, confidence in the effect estimates for critical outcomes. If one has high confidence for some critical outcomes (typically, benefits of an intervention), but low confidence for other outcomes considered critical (often long-term harms), then a weak recommendation is likely warranted. The more closely balanced the trade-offs between desirable and undesirable outcomes, the more likely that low confidence for any critical outcome will result in a weak recommendation.

Even when an apparently large gradient exists in the balance of desirable vs. undesirable outcomes, panels will be appropriately reluctant to offer a strong recommendation if their confidence in effect estimates is low. This is in part because when confidence in the estimate of effect is lower, choice is more preference dependent.

For instance, the GRADE approach provides insight into how guideline panels should have handled the decision regarding hormone replacement therapy (HRT) in postmenopausal women in the 1990s when observational studies suggested a substantial reduction in cardiovascular risk [24] (which randomized trials subsequently proved false [25], at least in women appreciably past the menopause), and equally low quality evidence suggested an increase in the risk of breast cancer (which proved true [26]).

Guideline panels during the 1990s made recommendations that were presented, or at least interpreted, as strong recommendations. Many primary care physicians, responding to these recommendations, enthusiastically encouraged their postmenopausal patients to use HRT. Appropriately considering the lack of confidence in estimates, women with a low level of risk aversion might indeed have been inclined to use HRT. Those with a high level of risk aversion would, however, have declined HRT. Clearly, a weak recommendation for (or perhaps even against) HRT would have been warranted.

For some questions, investigators may not have directly measured critical outcomes (in particular quality of life). In such instances, even if surrogates are available, confidence in estimates is very likely to be low.

2.3.1. Low confidence in effect estimates may, rarely, be tied to strong recommendations

In general, we discourage guideline panels from making strong recommendations when their confidence in estimates of effect for critical outcomes is low or very low. We have identified five paradigmatic situations, however, in which strong recommendations may be warranted despite low or very low quality of evidence (Table 4). These situations can be conceptualized as ones in which a panel would have a low level of regret if subsequent evidence showed that their recommendation was misguided.

One paradigmatic situation occurs when panels have low confidence regarding the benefit of an intervention in a life or death situation. Consider patients suffering from life-threatening disseminated blastomycosis [27]. High quality evidence suggests that amphotericin is more toxic than itraconazole, and low quality evidence that it reduces mortality in this context. When considering the subpopulation of patients with life-threatening blastomycosis, panels may reason that all or virtually all patients would choose the more toxic therapy given the very high risk of death and the possibility that amphotericin may decrease that risk. If they did so, they would make a strong recommendation for amphotericin.

In a second paradigmatic situation, panels may make a strong recommendation against an intervention when there is uncertainty of benefits, but they are confident about adverse effects and resource use. For example, it remains very uncertain whether whole-body computed tomography scan or magnetic resonance imaging screening confers benefits in terms of reduction of cancer risk, but there is no doubt that such tests generate false positives that result in anxiety and possibly invasive tests with their own discomfort and complications [28]. Such tests also consume scarce resources. Despite the low confidence with regard to benefits, guideline panels might legitimately make strong recommendations against screening imaging.

A third situation occurs when we have low quality evidence regarding relative benefit, but high quality evidence of lower harm for one of the competing alternatives. For instance, in patients who have early-stage, low-grade, *Helicobacter pylori*−positive gastric mucosa-associated lymphoid tissue lymphoma, low quality evidence suggests that initial *H. pylori* eradication therapy results in similar rates of complete response (50−80%) in comparison with the alternatives of radiation therapy or gastrectomy [29]. The evidence warrants high confidence in the increased morbidity associated with either radiation or gastrectomy vs. pharmacologic therapy. Furthermore, in patients without complete response, there is the option of later use of the higher risk alternatives. Thus, despite low confidence in estimates of effects, a strong recommendation for *H. pylori* eradication therapy appears appropriate.

In a fourth situation, panels may make strong recommendations for one of the two competing alternatives if they are confident of similarity of benefits, but have only

Table 4. Paradigmatic situations in which a strong recommendation may be warranted despite low or very low confidence in effect estimates

Situation	Condition	Example
1	When low quality evidence suggests benefit in a life-threatening situation (evidence regarding harms can be low or high)	Fresh frozen plasma or vitamin K in a patient receiving warfarin with elevated INR and an intracranial bleed. Only low quality evidence supports the benefits of limiting the extent of the bleeding
2	When low quality evidence suggests benefit and high quality evidence suggests harm or a very high cost	Head-to-toe CT/MRI screening for cancer. Low quality evidence of benefit of early detection but high quality evidence of possible harm and/or high cost (strong recommendation against this strategy)
3	When low quality evidence suggests equivalence of two alternatives, but high quality evidence of less harm for one of the competing alternatives	*Helicobacter pylori* eradication in patients with early stage gastric MALT lymphoma with *H. pylori* positive. Low quality evidence suggests that initial *H. pylori* eradication results in similar rates of complete response in comparison with the alternatives of radiation therapy or gastrectomy; high quality evidence suggests less harm/morbidity
4	When high quality evidence suggests equivalence of two alternatives and low quality evidence suggests harm in one alternative	Hypertension in women planning conception and in pregnancy. Strong recommendations for labetalol and nifedipine and strong recommendations against angiotensin converting enzyme (ACE) inhibitors and angiotensin receptor blockers (ARB)—all agents have high quality evidence of equivalent beneficial outcomes, with low quality evidence for greater adverse effects with ACE inhibitors and ARBs
5	When high quality evidence suggests modest benefits and low/very low quality evidence suggests possibility of catastrophic harm	Testosterone in males with or at risk of prostate cancer. High quality evidence for moderate benefits of testosterone treatment in men with symptomatic androgen deficiency to improve bone mineral density and muscle strength. Low quality evidence for harm in patients with or at risk of prostate cancer

Abbreviations: INR, international normalized ratio; CT, computed tomography; MRI, magnetic resonance imaging; MALT, mucosa-associated lymphoid tissue.

low or very low confidence regarding increased harm for one alternative. Reasoning that there is nothing to lose, and possibly a lot to gain in terms of a lower incidence of adverse effects, guideline panels may reasonably make a strong recommendation for the agent apparently free from serious toxicity. For instance, consider the management of hypertension in women who are planning conception and who are pregnant. There is high quality evidence of equivalent effectiveness for labetalol, nifedipine, angiotensin-converting enzyme (ACE) inhibitors, and angiotensin receptor blockers (ARBs). There is low quality evidence of harms for ACE inhibitors and ARBs. Panels have appropriately made strong recommendations for labetalol and nifedipine and strong recommendations against ACE inhibitors and ARBs [30].

A fifth paradigmatic situation occurs when we have moderate-to-high confidence about an intervention's modest benefits, but remain uncertain about its likelihood of causing catastrophic harm. For example, high quality evidence supports the inference that testosterone is beneficial for men with symptomatic androgen deficiency, improving their quality of life and markers of bone and muscle strength. However, low quality evidence links testosterone use to an increased risk of prostate cancer. As a result, a panel of endocrinologists formulated a strong recommendation against testosterone use in men with prostate cancer and in men pending evaluation of palpable prostate nodule or induration or prostate-specific antigen (PSA) level of 4 ng/mL or PSA level of 3 ng/mL in men at high risk of prostate cancer [31].

2.4. Resource use

Panels may or may not consider resource use in their judgments about the direction and strength of recommendations. Reasons for not considering resource use include a lack of reliable data, the intervention is not useful and the effort of calculating resource use can be spared, the desirable effects so greatly outweigh any undesirable effects that resource considerations would not alter the final judgment, or they have elected (or been instructed) to leave resource considerations up to other decision makers.

Once again, panels should be explicit about the decision they made not to consider resource utilization and the reason for their decision. If they elect to include resource utilization when making a recommendation, but have not included resource use as a consequence when preparing an evidence profile, they should be explicit about what types of resource use they considered when making the recommendation and whatever logic or evidence was used in their judgments.

For example, a panel making a recommendation about oseltamivir for treatment of patients hospitalized with avian influenza (H5N1) in nonpandemic situations considered the cost of oseltamivir, but did not explicitly consider the quality of the evidence for resource use. Overall, the quality of

the underlying evidence for all recommendations was rated as very low because it was based on small case series of H5N1 patients, on extrapolation from preclinical studies, and high quality studies of seasonal influenza. A strong recommendation to treat H5N1 patients with oseltamivir was made in part because of the severity of the disease. With only very low quality evidence of the beneficial and adverse effects of oseltamivir for avian influenza, the panel decided not to consider quality of evidence for resource use. The panel summarized their thinking regarding resource use as a factor in making their recommendation by stating: ''The cost is not high for treatment of sporadic cases'' [32].

We discuss special challenges related to rating the confidence in estimates for resource use in another article in this series [9].

3. Special considerations of the determinants of direction and strength of recommendations

3.1. Baseline risk (control event rate) can influence the balance

Table 3 presents an example of how guideline panels can move from evidence to recommendations in an explicit and transparent way. The final column in Table 3 presents the issues (if one calls the four determinants domains, then one might call these issues subdomains) that guideline panels should consider under each domain. One of these subdomains, which may be critical in the decision, is baseline risk.

Because, we usually determine absolute risk differences through applying the relative risk reduction to a baseline risk [11], large baseline risk differences will result in large absolute risk differences. For example, recommendations for duration of anticoagulation in patients with deep venous thrombosis will differ depending on the likelihood of recurrent thrombosis. The likelihood of recurrent thrombosis differs in those with and without clear precipitating factors for the original thrombotic event—in particular, patients whose deep venous thrombosis is precipitated by a surgical procedure have a low risk of recurrence. Anticoagulation is associated with inconvenience and a risk of serious bleeding. Therefore, indefinite anticoagulation will seldom be appropriate in those at low risk of recurrence whose absolute benefit with anticoagulation is small, but may well be mandated in patients at much higher risk. Thus, the strength of recommendations—and likely the direction—will differ in high- and low-risk groups [33].

3.2. Recommendations may differ by setting and perspective

In our introductory discussion of globalizing evidence, localizing recommendations, we noted that we do not expect uniformity of recommendations across settings. Here, we expand the reasons for the anticipated diversity, and how differences in perspective can contribute.

The impact of an intervention may differ across geographic settings depending on the risk of adverse events in untreated population (e.g., risk of coronary events is much lower in low income countries), or the capacity to deliver the intervention (e.g., monitoring of anticoagulant therapy).

Values and preferences may differ among cultures, even if those cultures appear very similar. For example, after viewing the same evidence, American and New Zealand guideline developers came to different conclusions about the trade-offs associated with colon cancer screening [34–36].

Values may also differ in subcultures vs. mainstream culture within a population. For example, in formulating the CCIRH guidelines, the panel's awareness of immigrant populations' vulnerability to family disruption and possible deportation supported the recommendation against routine screening for intimate partner violence [37].

Finally, resource implications and opportunity cost may differ. For instance, a year's supply of an expensive drug may cost the equivalent of a single nurse's salary in the United States, 4 nurses' salaries in Poland, and 20 nurses' salaries in China.

In the face of the same evidence, recommendations may also differ according to perspective. Our discussion in this article has addressed, almost exclusively, guideline panels making recommendations from the perspective of patients and the health care providers looking after those patients. Sometimes, however, a panel may make recommendations from a public health or societal perspective.

For example, panels making recommendations about H1N1, avian, or seasonal influenza may place a large value on outcomes that may not be directly critical or important to individual patients, such as reducing the spread of disease [32,38]. Other times, a panel may make recommendations from the perspective of the government or a private insurance company, placing a large value on costs (or alternative uses of resources) within a fixed budget. Equity, feasibility, and burden of illness may be other considerations important to public policy decision making, but of much less relevance to individual decision making. Panels should explicitly state the perspective they are taking, particularly when they are not taking a patient-centered perspective.

3.3. Evidence to recommendations synthesis

As in Table 3, GRADE suggests that guideline panels present a synthesis of their judgments about the domains determining direction and strength of recommendations, and how this synthesis informs the recommendation. Disagreement between panels is common [39–41], and disagreement may be a result of variability in judgments about the domains or of how panels synthesize those judgments. Presentation and publication of frameworks

summarizing the rationale for recommendations can support transparency in the decision process and be used for stakeholder engagement (Table 3).

Consider, for example, views expressed in the literature concerning the merits of perioperative use of beta-blockers in patients undergoing noncardiac surgery. Some assert that lower doses of beta-blockers administered well before surgery could prevent the documented increase in stroke risk with beta-blockers [42,43]. Others do not agree [44]. An evidence to action synthesis from the former group would emphasize the heterogeneity of results from trials that used different doses and different periods of administration of beta-blockers before surgery, and the latter would not.

Alternatively, disagreement in recommendations might be because they have different views of the relative value of reducing the risk of MI with beta-blocker use (approximately 1.5% in those at 5% baseline risk) vs. the increase in stroke risk (approximately 0.5% in those at 0.5% baseline risk of stroke). Both may agree that patients value preventing stroke more than preventing MI, but the synthesis from a panel recommending against beta-blockers would emphasize that the patients generally place very high value in avoiding disabling stroke and the asymptomatic nature of many perioperative MIs.

4. Conclusion

Patients, clinicians, and policy makers will all be better served by a more systematic and transparent system for judging the direction and strength of recommendations. Explicit presentation of how panels view the four domains to consider in the direction and strength of recommendations could play an important role in improving the transparency of panel decisions (Table 3).

References

[1] Guyatt G, Oxman A, Kunz R, Atkins D, Brozek J, Vist G, et al. GRADE guidelines: 2. Framing the question and deciding on important outcomes. J Clin Epidemiol 2011;64:395−400.

[2] Balshem H, Helfand M, Schunemann HJ, Oxman AD, Kunz R, Brozek J, et al. GRADE guidelines: 3. Rating the quality of evidence. J Clin Epidemiol 2011;64:401−6.

[3] Guyatt GH, Oxman AD, Vist G, Kunz R, Brozek J, Alonso-Coello P, et al. GRADE guidelines: 4. Rating the quality of evidence—study limitations (risk of bias). J Clin Epidemiol 2011;64:407−15.

[4] Guyatt GH, Oxman AD, Montori V, Vist G, Kunz R, Brozek J, et al. GRADE guidelines: 5. Rating the quality of evidence-publication bias. J Clin Epidemiol 2011;64:1277−82.

[5] Guyatt G, Oxman AD, Kunz R, Brozek J, Alonso-Coello P, Rind D, et al. GRADE guidelines 6. Rating the quality of evidence—imprecision. J Clin Epidemiol 2011;64:1283−93.

[6] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 7. Rating the quality of evidence-inconsistency. J Clin Epidemiol 2011;64:1294−302.

[7] Guyatt GH, Oxman AD, Kunz R, Woodcock J, Brozek J, Helfand M, et al. GRADE guidelines: 8. Rating the quality of evidence-indirectness. J Clin Epidemiol 2011;64:1303−10.

[8] Guyatt GH, Oxman AD, Sultan S, Glasziou P, Akl EA, Alonso-Coello P, et al. GRADE guidelines: 9. Rating up the quality of evidence. J Clin Epidemiol 2011;64:1311−6.

[9] Brunetti M, Shemilt I, Pregno S, Vale L, Oxman A, Lord J, et al. GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence. J Clin Epidemiol 2013;66:140−50.

[10] Guyatt GH, Oxman AD, Sultan S, Brozek J, Glasziou P, Alonso-Coello P, et al. GRADE guidelines: 11. Making an overall rating of confidence in effect estimates for a single outcome and for all outcomes. J Clin Epidemiol 2013;64:151−7.

[11] Guyatt GH, Oxman AD, Santesso N, Helfand M, Vist G, Kunz R, et al. GRADE guidelines: 12. Preparing summary of findings tables: binary outcomes. J Clin Epidemiol 2013;66:158−72.

[12] Guyatt GH, Thorlund K, Oxman AD, Walter S, Patrick D, Furukawa TA, et al. GRADE guidelines: 13. Preparing summary of findings tables: continuous outcomes. J Clin Epidemiol 2013;66:173−83.

[13] Guyatt G, Oxman A, Akl E, Kunz R, Vist G, Brozek J, et al. GRADE guidelines: 1. Introduction-GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383−94.

[14] Andrews J, Guyatt G, Oxman AD, Alderson P, Dahm P, Falck-Ytter Y, et al. GRADE guidelines: 14. Going from evidence to recommendations: the significance and presentation of recommendations. J Clin Epidemiol 2013;66:719−25.

[15] Eisenberg JM. Globalize the evidence, localize the decision: evidence-based medicine and international diversity. Health Aff (Millwood) 2002;21(3):166−8.

[16] Schunemann HJ, Woodhead M, Anzueto A, Buist S, Macnee W, Rabe KF, et al. A vision statement on guideline development for respiratory disease: the example of COPD. Lancet 2009;373:774−9.

[17] Montori V, Devereaux P, Straus S, Haynes B, Guyatt G. Decision making and the patient. In: Guyatt G, editor. The users' guides to the medical literature: a manual for evidence-based clinical practice. 2nd ed. New York, NY: McGraw-Hill; 2008.

[18] McLean S, Mulla S, Akl EA, Jankowski M, Vandvik P, Ibrahim S, et al. Patient values and preferences in decision making for antithrombotic therapy: a systematic review. Antithrombotic therapy and prevention of thrombosis, 9th ed.: American College of Chest physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141(2 Suppl):e1S−23S.

[19] Pottie K, Greenaway C, Feightner J, Welch V, Swinkels H, Rashid M, et al. Evidence-based clinical guidelines for immigrants and refugees. CMAJ 2011;183(12):E824−925.

[20] Bates S, Greer I, Middeldorp S, Veenstra D, Prabulos A, Vandvik P, et al. VTE, thrombophilia, antithrombotic therapy, and pregnancy: Antithrombotic therapy and prevention of thrombosis, 9th ed.: American College of Chest physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141(2 Suppl):e691S−736S.

[21] Goodman SG, Menon V, Cannon CP, Steg G, Ohman EM, Harrington RA. Acute ST-segment elevation myocardial infarction: American College of Chest physicians evidence-based clinical practice guidelines (8th Edition). Chest 2008;133(6 Suppl):708S−75S.

[22] Webster A, Woodroffe R, Taylor R, Chapman J, Craig J. Tacrolimus versus cyclosporin as primary immunosuppression for kidney transplant recipients. Cochrane Database Syst Rev 2006;4:CD003961.

[23] Devereaux PJ, Anderson DR, Gardner MJ, Putnam W, Flowerdew GJ, Brownell BF, et al. Differences between perspectives of physicians and patients on anticoagulation in patients with atrial fibrillation: observational study. BMJ 2001;323:1218−22.

[24] Guidelines for counseling postmenopausal women about preventive hormone therapy. American College of Physicians. Ann Intern Med 1992;117:1038−41.

[25] Hulley S, Grady D, Bush T, Furberg C, Herrington D, Riggs B, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women. Heart and Estrogen/progestin Replacement Study (HERS) Research Group. JAMA 1998;280:605−13.

[26] Rossouw JE, Anderson GL, Prentice RL, LaCroix AZ, Kooperberg C, Stefanick ML, et al. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. JAMA 2002;288: 321−33.

[27] Chapman SW, Dismukes WE, Proia LA, Bradsher RW, Pappas PG, Threlkeld MG, et al. Clinical practice guidelines for the management of blastomycosis: 2008 update by the Infectious Diseases Society of America. Clin Infect Dis 2008;46:1801−12.

[28] Lauenstein TC, Semelka RC. Emerging techniques: whole-body screening and staging with MRI. J Magn Reson Imaging 2006; 24(3):489−98.

[29] Malfertheiner P, Megraud F, O'Morain C, Bazzoli F, El-Omar E, Graham D, et al. Current concepts in the management of Helicobacter pylori infection: the Maastricht III Consensus Report. Gut 2007;56(6):772−81.

[30] Magee LA, Helewa M, Moutquin JM, van Dadelszen P, for the Hypertension Guideline Committee. Diagnosis, evaluation, and management of the hypertensive disorders of pregnancy. SOGC Clinical Practice Guideline, No. 206, March 2008. J Obstet Gynaecol Can 2008;30: S1−48.

[31] Bhasin S, Cunningham G, Hayes F, Matsumoto A, Snyder P, Swerdloff R, et al. Testosterone therapy in men with androgen deficiency syndromes: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2010;95:2536−59.

[32] Schunemann HJ, Hill SR, Kakad M, Bellamy R, Uyeki TM, Hayden FG, et al. WHO Rapid Advice Guidelines for pharmacological management of sporadic human infection with avian influenza A (H5N1) virus. Lancet Infect Dis 2007;7:21−31.

[33] Kearon C, Akl EA, Comerota AJ, Prandoni P, Bounameaux H, Goldhaber SZ, et al. Antithrombotic therapy for VTE disease: antithrombotic therapy and prevention of thrombosis, 9th ed.: American College of Chest physicians Evidence-Based Clinical Practice Guidelines. Chest 2012;141(2 Suppl):e419S−94S.

[34] Screening for colorectal cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 2008;149:627−37.

[35] Levin B, Lieberman DA, McFarland B, Andrews KS, Brooks D, Bond J, et al. Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology. Gastroenterology 2008;134(5):1570−95.

[36] Guidance on surveillance for people at increased risk of colorectal cancer. Wellington, New Zealand: New Zealand Guidelines Group; 2012.

[37] Tugwell P, Pottie K, Welch V, Ueffing E, Chambers A, Feightner J. Evaluation of evidence-based literature and formulation of recommendations for the clinical preventive guidelines for immigrants and refugees in Canada. CMAJ 2011;183(12):E933−8.

[38] Schunemann HJ, Hill SR, Kakad M, Vist GE, Bellamy R, Stockman L, et al. Transparent development of the WHO rapid advice guidelines. PLoS Med 2007;4(5):e119.

[39] Oxman AD, Glasziou P, Williams JW Jr. What should clinicians do when faced with conflicting recommendations? BMJ 2008;337:a2530.

[40] Georg G, Colombet I, Durieux P, Menard J, Meneton P. A comparative analysis of four clinical guidelines for hypertension management. J Hum Hypertens 2008;22(12):829−37.

[41] Matthys J, De Meyere M, van Driel ML, De Sutter A. Differences among international pharyngitis guidelines: not just academic. Ann Fam Med 2007;5(5):436−43.

[42] Kaafarani HM, Atluri PV, Thornby J, Itani KM. beta-Blockade in noncardiac surgery: outcome at all levels of cardiac risk. Arch Surg 2008;143:940−4. discussion 944.

[43] van Lier F, Schouten O, van Domburg RT, van der Geest PJ, Boersma E, Fleisher LA, et al. Effect of chronic beta-blocker use on stroke after noncardiac surgery. Am J Cardiol 2009;104:429−33.

[44] Bangalore S, Wetterslev J, Pranesh S, Sawhney S, Gluud C, Messerli FH. Perioperative beta blockers in patients having noncardiac surgery: a meta-analysis. Lancet 2008;372:1962−76.

The BMJ
jblock@bmj.com
Cite this as: *BMJ* 2023;380:p382
http://dx.doi.org/10.1136/bmj.p382
Published: 23 February 2023

BMJ INVESTIGATION

Gender dysphoria in young people is rising—and so is professional disagreement

More children and adolescents are identifying as transgender and are being offered medical treatment, especially in the US—but some providers and European authorities are urging caution because of a lack of strong evidence. Jennifer Block reports

Jennifer Block *investigations reporter*

Last October the American Academy of Pediatrics (AAP) gathered inside the Anaheim Convention Center in California for its annual conference. Outside, several dozen people rallied to hear speakers including Abigail Martinez, a mother whose child began hormone treatment at age 16 and died by suicide at age 19. Supporters chanted the teen's given name, Yaeli; counter protesters chanted, "Protect trans youth!" For viewers on a livestream, the feed was interrupted as the two groups fought for the camera.

The AAP conference is one of many flashpoints in the contentious debate in the United States over if, when, and how children and adolescents with gender dysphoria should be medically or surgically treated. US medical professional groups are aligned in support of "gender affirming care" for gender dysphoria, which may include gonadotrophin releasing hormone analogues (GnRHa) to suppress puberty; oestrogen or testosterone to promote secondary sex characteristics; and surgical removal or augmentation of breasts, genitals, or other physical features. At the same time, however, several European countries have issued guidance to limit medical intervention in minors, prioritising psychological care.

The discourse is polarised in the US. Conservative politicians, pundits, and social media influencers accuse providers of pushing "gender ideology" and even "child abuse," lobbying for laws banning medical transition for minors. Progressives argue that denying access to care is a transphobic violation of human rights. There's little dispute within the medical community that children in distress need care, but concerns about the rapid widespread adoption of interventions and calls for rigorous scientific review are coming from across the ideological spectrum.[1]

The surge in treatment of minors

More adolescents with no history of gender dysphoria—predominantly birth registered females[2]—are presenting at gender clinics. A recent analysis of insurance claims by Komodo Health found that nearly 18 000 US minors began taking puberty blockers or hormones from 2017 to 2021, the number rising each year.[3 4] Surveys aiming to measure prevalence have found that about 2% of high school aged teens identify as "transgender."[5] These young people are also more likely than their cisgender peers to have concurrent mental health and neurodiverse conditions including depression, anxiety, attention deficit disorders, and autism.[6] In the US, although Medicaid coverage varies by state and by treatment, the Biden administration has warned states that not covering care is in violation of federal law prohibiting discrimination.[7] Meanwhile, the number of private clinics that focus on providing hormones and surgeries has grown from just a few a decade ago to more than 100 today.[4]

As the number of young people receiving medical transition treatments rises, so have the voices of those who call themselves "detransitioners" or "retransitioners," some of whom claim that early treatment caused preventable harm.[8] Large scale, long term research is lacking,[9] and researchers disagree about how to measure the phenomenon, but two recent studies suggest that as many as 20-30% of patients may discontinue hormone treatment within a few years.[10 11] The World Professional Association for Transgender Health (WPATH) asserts that detransition is "rare."[12]

Chloe Cole, now aged 18, had a double mastectomy at age 15 and spoke at the AAP rally. "Many of us were young teenagers when we decided, on the direction of medical experts, to pursue irreversible hormone treatments and surgeries," she read from her tablet at the rally, which had by this time moved indoors to avoid confrontation. "This is not informed consent but a decision forced under extreme duress."

Scott Hadland, chief of adolescent medicine at Massachusetts General Hospital and Harvard Medical School, dismissed the "handful of cruel protesters" outside the AAP meeting in a tweet that morning. He wrote, "Inside 10 000 pediatricians stand in solidarity for trans & gender diverse kids & their families to receive evidence-based, lifesaving, individualized care."[13]

Same evidence, divergent recommendations

Three organisations have had a major role in shaping the US's approach to gender dysphoria care: WPATH, the AAP, and the Endocrine Society (see box). On 15 September 2022 WPATH published the eighth edition of its Standards of Care for the Health of Transgender and Gender Diverse People, with new chapters on children and adolescents and no minimum age requirements for hormonal and surgical treatments.[2 12] GnRHa treatment, says WPATH, can

be initiated to arrest puberty at its earliest stage, known as Tanner stage 2.

The Endocrine Society also supports hormonal and surgical intervention in adolescents who meet criteria in clinical practice guidelines published in 2009 and updated in 2017.[14] And the AAP's 2018 policy statement, *Ensuring Comprehensive Care and Support for Transgender and Gender-Diverse Children and Adolescents*, says that "various interventions may be considered to better align" a young person's "gender expression with their underlying identity."[15] Among the components of "gender affirmation" the AAP names social transition, puberty blockers, sex hormones, and surgeries. Other prominent professional organisations, such as the American Medical Association, have issued policy statements in opposition to legislation that would curtail access to medical treatment for minors.[16 -19]

These documents are often cited to suggest that medical treatment is both uncontroversial and backed by rigorous science. "All of those medical societies find such care to be evidence-based and medically necessary," stated a recent article on transgender healthcare for children published in *Scientific American*.[20] "Transition related healthcare is not controversial in the medical field," wrote Gillian Branstetter, a frequent spokesperson on transgender issues currently with the American Civil Liberties Union, in a 2019 guide for reporters.[21] Two physicians and an attorney from Yale recently opined in the *Los Angeles Times* that "gender-affirming care is standard medical care, supported by major medical organizations . . . Years of study and scientific scrutiny have established safe, evidence-based guidelines for delivery of lifesaving, gender-affirming care."[22] Rachel Levine, the US assistant secretary for health, told National Public Radio last year regarding such treatment, "There is no argument among medical professionals."[23]

Internationally, however, governing bodies have come to different conclusions regarding the safety and efficacy of medically treating gender dysphoria. Sweden's National Board of Health and Welfare, which sets guidelines for care, determined last year that the risks of puberty blockers and treatment with hormones "currently outweigh the possible benefits" for minors.[24] Finland's Council for Choices in Health Care, a monitoring agency for the country's public health services, issued similar guidelines, calling for psychosocial support as the first line treatment.[25] (Both countries restrict surgery to adults.)

Medical societies in France, Australia, and New Zealand have also leant away from early medicalisation.[26 27] And NHS England, which is in the midst of an independent review of gender identity services, recently said that there was "scarce and inconclusive evidence to support clinical decision making"[28] for minors with gender dysphoria[29] and that for most who present before puberty it will be a "transient phase," requiring clinicians to focus on psychological support and to be "mindful" even of the risks of social transition.[30]

Box: The origins of paediatric gender medicine in the United States

The World Professional Association for Transgender Health (WPATH) began as a US based advocacy group and issued the first edition of the Standards of Care in 1979, when it was serving a small population of mostly adult male-to-female transsexuals. "WPATH became the standard because there was nobody else doing it," says Erica Anderson, a California based clinical psychologist and former WPATH board member. The professional US organisations that lined up in support "looked heavily to WPATH and the Endocrine Society for their guidance," she told *The BMJ*.

The Endocrine Society's guidance for adolescents grew out of clinicians' research in the Netherlands in the late 1990s and early 2000s. Peggy Cohen-Kettenis, a Utrecht gender clinic psychologist, collaborated with endocrinologists in Amsterdam, one of whom had experience of prescribing gonadotrophin releasing hormone analogues, relatively new at the time. Back then, gender dysphoric teens had to wait until the age of majority for sex hormones, but the team proposed that earlier intervention could benefit carefully selected minors.[40]

The clinic treated one natal female patient with triptorelin, published a case study and feasibility proposal, and began treating a small number of children at the turn of the millennium. The Dutch Protocol was published in 2006, referring to 54 children whose puberty was being suppressed and reporting preliminary results on the first 21.[41] The researchers received funding from Ferring Pharmaceuticals, the manufacturer of triptorelin.

In 2007 the endocrinologist Norman Spack began using the protocol at Boston Children's Hospital and joined Cohen-Kettenis and her Dutch colleagues in writing the Endocrine Society's first clinical practice guideline.[42] When that was published in 2009, puberty had been suppressed in just over 100 gender dysphoric young people.[40]

American Academy of Pediatrics (AAP) committee members began discussing the need for a statement in 2014, four years before publication, says Jason Rafferty, assistant professor of paediatrics and psychiatry at Brown University, Rhode Island, and the statement's lead author. "The AAP recognised that it had a responsibility to provide some clinical guidance, but more importantly to come out with a statement that said we need research, we need to integrate the principles of gender affirmative care into medical education and into child health," he says. "What our policy statement is not meant to be is a protocol or guidelines in and of themselves."

"Don't call them evidence based"

"The brief history of guidelines is that, going back more than 30 years ago, experts would write articles and so on about what people should do. But formal guidelines as we think of them now were seldom or non-existent," says Gordon Guyatt, distinguished professor in the Department of Health Research Methods, Evidence, and Impact at McMaster University, Ontario.

That led to the movement towards developing criteria for what makes a "trustworthy guideline," of which Guyatt was a part.[31] One pillar of this, he told *The BMJ*, is that they "are based on systematic review of the relevant evidence," for which there are also now standards, as opposed to a traditional narrative literature review in which "a bunch of experts write whatever they felt like using no particular standards and no particular structure."

Mark Helfand, professor of medical informatics and clinical epidemiology at Oregon Health and Science University, says, "An evidence based recommendation requires two steps." First, "an unbiased, thorough, critical systematic review of all the relevant evidence." Second, "some commitment to link the strength of the recommendations to the quality of the evidence."

The Endocrine Society commissioned two systematic reviews for its clinical practice guideline, *Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons*: one on the effects of sex steroids on lipids and cardiovascular outcomes, the other on their effects on bone health.[32 33] To indicate the quality of evidence underpinning its various guidelines, the Endocrine Society employed the GRADE system (grading of recommendations assessment, development, and evaluation) and judged the quality of evidence for all recommendations on adolescents as "low" or "very low."

Guyatt, who co-developed GRADE, found "serious problems" with the Endocrine Society guidelines, noting that the systematic reviews

didn't look at the effect of the interventions on gender dysphoria itself, arguably "the most important outcome." He also noted that the Endocrine Society had at times paired strong recommendations—phrased as "we recommend"—with weak evidence. In the adolescent section, the weaker phrasing "we suggest" is used for pubertal hormone suppression when children "first exhibit physical changes of puberty"; however, the stronger phrasing is used to "recommend" GnRHa treatment.

"GRADE discourages strong recommendations with low or very low quality evidence except under very specific circumstances," Guyatt told *The BMJ*. Those exceptions are "very few and far between," and when used in guidance, their rationale should be made explicit, Guyatt said. In an emailed response, the Endocrine Society referenced the GRADE system's five exceptions, but did not specify which it was applying.

Helfand examined the recently updated WPATH Standards of Care and noted that it "incorporated elements of an evidence based guideline." For one, WPATH commissioned a team at Johns Hopkins University in Maryland to conduct systematic reviews.[34 35] However, WPATH's recommendations lack a grading system to indicate the quality of the evidence—one of several deficiencies. Both Guyatt and Helfand noted that a trustworthy guideline would be transparent about all commissioned systematic reviews: how many were done and what the results were. But Helfand remarked that neither was made clear in the WPATH guidelines and also noted several instances in which the strength of evidence presented to justify a recommendation was "at odds with what their own systematic reviewers found."

For example, one of the commissioned systematic reviews found that the strength of evidence for the conclusions that hormonal treatment "may improve" quality of life, depression, and anxiety among transgender people was "low," and it emphasised the need for more research, "especially among adolescents."[35] The reviewers also concluded that "it was impossible to draw conclusions about the effects of hormone therapy" on death by suicide.

Despite this, WPATH recommends that young people have access to treatments after comprehensive assessment, stating that the "emerging evidence base indicates a general improvement in the lives of transgender adolescents."[12] And more globally, WPATH asserts, "There is strong evidence demonstrating the benefits in quality of life and well-being of gender-affirming treatments, including endocrine and surgical procedures," procedures that "are based on decades of clinical experience and research; therefore, they are not considered experimental, cosmetic, or for the mere convenience of a patient. They are safe and effective at reducing gender incongruence and gender dysphoria."[12]

Those two statements are each followed by more than 20 references, among them the commissioned systematic review. This stood out to Helfand as obscuring which conclusions were based on evidence versus opinion. He says, "It's a very strange thing to feel that they had to cite some of the studies that would have been in the systematic review or purposefully weren't included in the review, because that's what the review is for."

For minors, WPATH contends that the evidence is so limited that "a systematic review regarding outcomes of treatment in adolescents is not possible." But Guyatt counters that "systematic reviews are always possible," even if few or no studies meet the eligibility criteria. If an entity has made a recommendation without one, he says, "they'd be violating standards of trustworthy guidelines." Jason Rafferty, assistant professor of paediatrics and psychiatry at Brown University, Rhode Island, and lead author of the AAP statement, remarks that the AAP's process "doesn't quite fit the definition of systematic review, but it is very comprehensive."

Sweden conducted systematic reviews in 2015 and 2022 and found the evidence on hormonal treatment in adolescents "insufficient and inconclusive."[24] Its new guidelines note the importance of factoring the possibility that young people will detransition, in which case "gender confirming treatment thus may lead to a deteriorating of health and quality of life (i.e., harm)."

Cochrane, an international organisation that has built its reputation on delivering independent evidence reviews, has yet to publish a systematic review of gender treatments in minors. But *The BMJ* has learnt that in 2020 Cochrane accepted a proposal to review puberty blockers and that it worked with a team of researchers through 2021 in developing a protocol, but it ultimately rejected it after peer review. A spokesperson for Cochrane told *The BMJ* that its editors have to consider whether a review "would add value to the existing evidence base," highlighting the work of the UK's National Institute for Health and Care Excellence, which looked at puberty blockers and hormones for adolescents in 2021. "That review found the evidence to be inconclusive, and there have been no significant primary studies published since."

In 2022 the state of Florida's Agency for Health Care Administration commissioned an overview of systematic reviews looking at outcomes "important to patients" with gender dysphoria, including mental health, quality of life, and complications. Two health research methodologists at McMaster University carried out the work, analysing 61 systematic reviews and concluding that "there is great uncertainty about the effects of puberty blockers, cross-sex hormones, and surgeries in young people." The body of evidence, they said, was "not sufficient" to support treatment decisions.

Calling a treatment recommendation "evidence based" should mean that a treatment has not just been systematically studied, says Helfand, but that there was also a finding of high quality evidence supporting its use. Weak evidence "doesn't just mean something esoteric about study design, it means there's uncertainty about whether the long term benefits outweigh the harms," Helfand adds.

"Evidence itself never tells you what to do," says Guyatt. That's why guidelines must make explicit the values and preferences that underlie the recommendation.

The Endocrine Society acknowledges in its recommendations on early puberty suppression that it is placing "a high value on avoiding an unsatisfactory physical outcome when secondary sex characteristics have become manifest and irreversible, a higher value on psychological well-being, and a lower value on avoiding potential harm."[14]

WPATH acknowledges that while its latest guidelines are "based upon a more rigorous and methodological evidence-based approach than previous versions," the evidence "is not only based on the published literature (direct as well as background evidence) but also on consensus-based expert opinion." In the absence of high quality evidence and the presence of a patient population in need—who are willing to take on more personal risk—consensus based guidelines are not unwarranted, says Helfand. "But don't call them evidence based."

An evidence base under construction

In 2015 the US National Institutes of Health awarded a $5.7m (£4.7m; €5.3m) grant to study "the impact of early medical treatment in transgender youth."[36] The abstract submitted by applicants said that the study was "the first in the US to evaluate longitudinal

outcomes of medical treatment for transgender youth and will provide essential evidence-based data on the physiological and psychosocial effects and safety" of current treatments. Researchers are following two groups, one of participants who began receiving GnRHa in early puberty and another group who began cross sex hormone treatment in adolescence. The study doesn't include a concurrent no-treatment control group.

Robert Garofalo, chief of adolescent medicine at the Lurie Children's Hospital in Chicago and one of four principal investigators, told a podcast interviewer in May 2022 that the evidence base remained "a challenge . . . it is a discipline where the evidence base is now being assembled" and that "it's truly lagging behind [clinical practice], I think, in some ways." That care, he explained, was "being done safely. But only now, I think, are we really beginning to do the type of research where we're looking at short, medium, and long term outcomes of the care that we are providing in a way that I think hopefully will be either reassuring to institutions and families and patients or also will shed a light on things that we can be doing better."[37]

While Garofalo was doing the research he served as "contributor" on the AAP's widely cited 2018 policy statement, which recommends that children and adolescents "have access to comprehensive, gender-affirming, and developmentally appropriate health care," including puberty blockers, sex hormones, and, on a case-by-case basis, surgeries.[15]

Garofalo said in the May interview, "There is universal support for gender affirming care from every mainstream US based medical society that I can think of: the AMA, the APA, the AAP. I mean, these organisations never agree with one another." Garofalo declined an interview and did not respond to *The BMJ*'s requests for comment.

The rush to affirm

Sarah Palmer, a paediatrician in private practice in Indiana, is one of five coauthors of a 2022 resolution submitted to the AAP's leadership conference asking that it revisit the policy after "a rigorous systematic review of available evidence regarding the safety, efficacy, and risks of childhood social transition, puberty blockers, cross sex hormones and surgery." In practice, Palmer told *The BMJ*, clinicians define "gender affirming" care so broadly that "it's been taken by many people to mean go ahead and do anything that affirms. One of the main things I've seen it used for is masculinising chest surgery, also known as mastectomy in teenage patients." The AAP has told *The BMJ* that all policy statements are reviewed after five years and so a "revision is under way," based on its experts' own "robust evidence review."

Palmer says, "I've seen a quick evolution, from kids with a very rare case of gender dysphoria who were treated with a long course of counselling and exploration before hormones were started," to treatment progressing "very quickly—even at the first visit to gender clinic—and there's no psychologist involved anymore."

Laura Edwards-Leeper, a clinical psychologist who worked with the endocrinologist Norman Spack in Boston and coauthored the WPATH guidelines for adolescents, has observed a similar trend. "More providers do not value the mental health component," she says, so in some clinics families come in and their child is "pretty much fast tracked to medical intervention." In a study of teens at Seattle Children's Hospital's gender clinic, two thirds were taking hormones within 12 months of the initial visit.[38]

The British paediatrician Hilary Cass, in her interim report of a UK review into services for young people with gender identity issues, noted that some NHS staff reported feeling "under pressure to adopt an unquestioning affirmative approach and that this is at odds with the standard process of clinical assessment and diagnosis that they have been trained to undertake in all other clinical encounters."

Eli Coleman, lead author of WPATH's Standards of Care and former director of the Institute for Sexual and Gender Health at the University of Minnesota, told *The BMJ* that the new guidelines emphasised "careful assessment prior to any of these interventions" by clinicians who have appropriate training and competence to assure that minors have "the emotional and cognitive maturity to understand the risks and benefits." He adds, "What we know and what we don't know has to be explained to youth and their parents or caregivers in a balanced way which really details that this is the evidence that we have, that we obviously would like to have more evidence, and that this is a risk-benefit scenario that you have to consider."

Joshua Safer, director of the Center for Transgender Medicine and Surgery at Mount Sinai Hospital in New York and coauthor of the Endocrine Society guidelines, told *The BMJ* that assessment is standard practice at the programme he leads. "We start with a mental health evaluation for anybody under the age of 18," he says. "There's a lot of talking going on—that's a substantial element of things." Safer has heard stories of adolescents leaving a first or second appointment with a prescription in hand but says that these are overblown. "We really do screen these kids pretty well, and the overwhelming majority of kids who get into these programmes do go on to other interventions," he says.

Without an objective diagnostic test, however, others remain concerned. The demand for services has led to a "perfunctory informed consent process," wrote two clinicians and a researcher in a recent issue of the *Journal of Sex and Marital Therapy*,[39] in spite of two key uncertainties: the long term impacts of treatment and whether a young person will persist in their gender identity. And the widespread impression of medical consensus doesn't help. "Unfortunately, gender specialists are frequently unfamiliar with, or discount the significance of, the research in support of these two concepts," they wrote. "As a result, the informed consent process rarely adequately discloses this information to patients and their families."

For Guyatt, claims of certainty represent both the success and failure of the evidence based medicine movement. "Everybody now has to claim to be evidence based" in order to be taken seriously, he says—that's the success. But people "don't particularly adhere to the standard of what is evidence based medicine—that's the failure." When there's been a rigorous systematic review of the evidence and the bottom line is that "we don't know,'" he says, then "anybody who then claims they do know is not being evidence based."

This feature has been funded by the BMJ Investigations Unit. For details see bmj.com/investigations.

Competing interests: I have read and understood BMJ policy on declaration of interests and have no relevant interests to declare.

Provenance: commissioned; externally peer reviewed.

1 Parker K, Horowitz JM, Brown A. Americans' complex views on gender identity and transgender issues. Pew Research Center's Social & Demographic Trends Project. 2022. https://www.pewresearch.org/social-trends/2022/06/28/americans-complex-views-on-gender-identity-and-transgender-issues/

2 Block J. US transgender health guidelines leave age of treatment initiation open to clinical judgment. *BMJ* 2022;378:. doi: 10.1136/bmj.o2303 pmid: 36167353

3 Respaut R, Terhune C. Number of transgender children seeking treatment surges in US. *Reuters* 2022 Oct 6. https://www.reuters.com/investigates/special-report/usa-transyouth-data/

4 Terhune C, Respaut R, Conlin M. As children line up at gender clinics, families confront many unknowns. *Reuters* 2022 Oct 6. https://www.reuters.com/investigates/special-report/usa-transyouth-care/

5 Johns MM, Lowry R, Andrzejewski J, etal. Transgender identity and experiences of violence victimization, substance use, suicide risk, and sexual risk behaviors among high school students—19 States and large urban school districts, 2017. *MMWR Morb Mortal Wkly Rep* 2019;68:-71. . doi: 10.15585/mmwr.mm6803a3 pmid: 30677012

6 Becerra-Culqui TA, Liu Y, Nash R, etal. Mental health of transgender and gender nonconforming youth compared with their peers. *Pediatrics* 2018;141:e20173845. doi: 10.1542/peds.2017-3845 pmid: 29661941

7 Gomez I, Ranji U, Salganicoff A, et al. Update on Medicaid coverage of gender-affirming health services. KFF. 2022 https://www.kff.org/womens-health-policy/issue-brief/update-on-medicaid-coverage-of-gender-affirming-health-services/

8 Littman L. Individuals treated for gender dysphoria with medical and/or surgical transition who subsequently detransitioned: a survey of 100 detransitioners. *Arch Sex Behav* 2021;50:-69. doi: 10.1007/s10508-021-02163-w pmid: 34665380

9 Respaut R, Terhune C, Conlin M. Why detransitioners are crucial to the science of gender care. *Reuters* 2022 Dec 22. https://www.reuters.com/investigates/special-report/usa-transyouth-outcomes/

10 Boyd I, Hackett T, Bewley S. Care of Transgender Patients: A General Practice Quality Improvement Approach. *Healthcare (Basel)* 2022;10:. doi: 10.3390/healthcare10010121. pmid: 35052285

11 Roberts CM, Klein DA, Adirim TA, Schvey NA, Hisle-Gorman E. Continuation of Gender-affirming Hormones Among Transgender Adolescents and Adults. *J Clin Endocrinol Metab* 2022;107:-43. doi: 10.1210/clinem/dgac251. pmid: 35452119

12 Coleman E, Radix AE, Bouman WP, etal. Standards of care for the health of transgender and gender diverse people, version 8. *Int J Transgend Health* 2022;23(Suppl 1):-259. doi: 10.1080/26895269.2022.2100644 pmid: 36238954

13 Hadland S. A handful of cruel protesters outside the #AAP2022 meeting, but inside 10 000 pediatricians stand in solidarity for trans & gender-diverse kids & their families to receive evidence based, lifesaving, individualized care between patients, parents & their doctor. @AmerAcadPeds: pic.twitter.com/b2K2JdRnX9. Twitter. 2022. https://twitter.com/DrScottHadland/status/1578815082590400512

14 Hembree WC, Cohen-Kettenis PT, Gooren L, etal. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab* 2017;102:-903. doi: 10.1210/jc.2017-01658 pmid: 28945902

15 Rafferty JCommittee on Psychosocial Aspects of Child and Family HealthCommittee on AdolescenceSection on Lesbian, Gay, Bisexual, and Transgender Health and Wellness. Ensuring comprehensive care and support for transgender and gender-diverse children and adolescents. *Pediatrics* 2018;142:e20182162. doi: 10.1542/peds.2018-2162 pmid: 30224363

16 American Academy of Child and Adolescent Psychiatry. AACAP statement responding to efforts to ban evidence-based care for transgender and gender diverse. https://www.aacap.org/AACAP/Latest_News/AACAP_Statement_Responding_to_Efforts-to_ban_Evidence-Based_Care_for_Transgender_and_Gender_Diverse.aspx

17 American Medical Association. March 26, 2021: State advocacy update. 2021. https://www.ama-assn.org/health-care-advocacy/advocacy-update/march-26-2021-state-advocacy-update

18 American Psychological Association. Resolution on supporting sexual/gender diverse children and adolescents in schools. 2020. https://www.apa.org/pi/lgbt/resources/policy/gender-diverse-children

19 American Psychiatric Association. Position statement on treatment of transgender (trans) and gender diverse youth. 2020. https://www.psychiatry.org/File%20Library/About-APA/Organization-Documents-Policies/Policies/Position-Transgender-Gender-Diverse-Youth.pdf

20 Boerner H. What the science on gender-affirming care for transgender kids really shows. *Sci Am* 2022 (published online 12 May). https://www.scientificamerican.com/article/what-the-science-on-gender-affirming-care-for-transgender-kids-really-shows/

21 Branstetter G. Transgender youth & health care: a guide for reporters. *Medium* 2019. https://gillbranstetter.medium.com/transgender-youth-health-care-a-guide-for-reporters-820f8fbaff21

22 Olezeski C, McNamara M, Alstott A. Op-ed: Denying trans youth gender-affirming care is an affront to science and medical ethics. *Los Angeles Times* 2022. https://www.latimes.com/opinion/story/2022-06-13/trans-youth-healthcare-state-bans

23 Simmons-Duffin S. Rachel Levine calls state anti-LGBTQ bills disturbing and dangerous to trans youth. *NPR* 2022 Apr 29. https://www.npr.org/sections/health-shots/2022/04/29/1095227346/rachel-levine-calls-state-anti-lgbtq-bills-disturbing-and-dangerous-to-trans-you

24 Socialstyrelsen: National Board of Health and Welfare. Care of children and adolescents with gender dysphoria. Report 2022-3-7799. 2022. https://www.socialstyrelsen.se/globalassets/sharepoint-dokument/artikelkatalog/kunskapsstod/2022-3-7799.pdf

25 Palveluvalikoima (Council for Choices in Health Care in Finland). Medical treatment methods for gender dysphoria in non-binary adults—recommendation. Jun 2020. https://palveluvalikoima.fi/documents/1237350/22895623/Summary_non-binary_en.pdf/8e5f9035-6c98-40d9-6acd-7459516d6f92/Summary_non-binary_en.pdf?t=1592318035000

26 Académie Nationale de Médecine. La médecine face à la transidentité de genre chez les enfants et les adolescents [Medicine and gender transidentity in children and adolescents. 25 Feb 2022. https://www.academie-medecine.fr/la-medecine-face-a-la-transidentite-de-genre-chez-les-enfants-et-les-adolescents/?lang=en (In French and English)

27 Royal Australian and New Zealand College of Psychiatrists. Recognising and addressing the mental health needs of people experiencing gender dysphoria/gender incongruence. 2021. https://www.ranzcp.org/news-policy/policy-and-advocacy/position-statements/gender-dysphoria

28 NHS England. Implementing advice from the Cass review. https://www.england.nhs.uk/commissioning/spec-services/npc-crg/gender-dysphoria-clinical-programme/implementing-advice-from-the-cass-review/

29 Cass Review. NICE evidence reviews. 2021. https://cass.independent-review.uk/nice-evidence-reviews/

30 NHS England. Interim service specification: specialist service for children and young people with gender dysphoria (phase 1 providers). Oct 2022. https://www.engage.england.nhs.uk/specialised-commissioning/gender-dysphoria-services/user_uploads/b1937-ii-specialist-service-for-children-and-young-people-with-gender-dysphoria-1.pdf

31 Institute of MedicineBoard on Health Care ServicesCommittee on Standards for Developing Trustworthy Clinical Practice Guidelines. *Clinical practice guidelines we can trust* . National Academies Press, 2011. https://play.google.com/store/books/details?id=b_RTRs8SEOYC

32 Maraka S, Singh Ospina N, Rodriguez-Gutierrez R, etal. Sex steroids and cardiovascular outcomes in transgender individuals: a systematic review and meta-analysis. *J Clin Endocrinol Metab* 2017;102:-23. doi: 10.1210/jc.2017-01643 pmid: 28945852

33 Singh-Ospina N, Maraka S, Rodriguez-Gutierrez R, etal. Effect of sex steroids on the bone health of transgender individuals: a systematic review and meta-analysis. *J Clin Endocrinol Metab* 2017;102:-13. doi: 10.1210/jc.2017-01642 pmid: 28945851

34 Wilson LM, Baker KE, Sharma R, Dukhanin V, McArthur K, Robinson KA. Effects of antiandrogens on prolactin levels among transgender women on estrogen therapy: A systematic review. *Int J Transgend Health* 2020;21:-402. doi: 10.1080/15532739.2020.1819505 pmid: 34993517

35 Baker KE, Wilson LM, Sharma R, Dukhanin V, McArthur K, Robinson KA. Hormone therapy, mental health, and quality of life among transgender people: a systematic review. *J Endocr Soc* 2021;5:bvab011.

36 National Institutes of Health. The impact of early medical treatment in transgender youth. NIH Reporter. https://reporter.nih.gov/project-details/8965408

37 Northwestern University Feinberg School of Medicine. Evidence-based gender-affirming care for young adults with Robert Garofalo, MD, MPH. 20 May 2022. https://www.feinberg.northwestern.edu/research/news/podcast/2022/evidence-based-gender-affirming%20-care-for-young-adults-with-robert-Garofalo-md-mph.html

38 Tordoff DM, Wanta JW, Collin A, Stepney C, Inwards-Breland DJ, Ahrens K. Mental health outcomes in transgender and nonbinary youths receiving gender-affirming care. *JAMA Netw Open* 2022;5:e220978. doi: 10.1001/jamanetworkopen.2022.0978 pmid: 35212746

39 Levine SB, Abbruzzese E, Mason JW. Reconsidering informed consent for trans-identified children, adolescents, and young adults. *J Sex Marital Ther* 2022;48:-27. doi: 10.1080/0092623X.2022.2046221 pmid: 35300570

40 Biggs M. The Dutch Protocol for Juvenile Transsexuals: Origins and Evidence. *J Sex Marital Ther* 2022:-21. doi: 10.1080/0092623X.2022.2121238 pmid: 36120756

41 Delemarre-van de Waal HA, Cohen-Kettenis PT. Clinical management of gender identity disorder in adolescents: a protocol on psychological and paediatric endocrinology aspects. *Eur J Endocrinol* 2006;155(Supp 1):-7. https://academic.oup.com/ejendo/article/155/Supplement_1/S131/6695708doi: 10.1530/eje.1.02231.

42 Hembree WC, Cohen-Kettenis P, Delemarre-van de Waal HA, etalEndocrine Society. Endocrine treatment of transsexual persons: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab* 2009;94:-54. doi: 10.1210/jc.2009-0345 pmid: 19509099

Journal of the Endocrine Society, 2021, Vol. 5, No. 4, 1–16
doi:10.1210/jendso/bvab011
Meta-Analysis

Meta-Analysis

Hormone Therapy, Mental Health, and Quality of Life Among Transgender People: A Systematic Review

Kellan E. Baker,[1,3] Lisa M. Wilson,[1,3] Ritu Sharma,[1,3] Vadim Dukhanin,[1,3] Kristen McArthur,[1,3] and Karen A. Robinson[2,3]

[1]Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD 21205, USA; [2]Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA; and [3]Johns Hopkins Evidence-Based Practice Center, 615 North Wolfe Street, Baltimore, MD 21205, USA

ORCiD number: 0000-0002-9716-7936 (K. E. Baker).

Abbreviations: BDI, Beck Depression Inventory; ENIGI, European Network for the Investigation of Gender Incongruence; GnRH, gonadotropin-releasing hormone; HADS, Hospital Anxiety and Depression Scale; QOL, quality of life; RCT, randomized controlled trial; SF-36, Short Form-36 Health Survey; WPATH, World Professional Association for Transgender Health.

Received: 5 October 2020; Editorial Decision: 25 January 2021; First Published Online: 2 February 2021; Corrected and Typeset: 19 February 2021.

Abstract

We sought to systematically review the effect of gender-affirming hormone therapy on psychological outcomes among transgender people. We searched PubMed, Embase, and PsycINFO through June 10, 2020 for studies evaluating quality of life (QOL), depression, anxiety, and death by suicide in the context of gender-affirming hormone therapy among transgender people of any age. We excluded case studies and studies reporting on less than 3 months of follow-up. We included 20 studies reported in 22 publications. Fifteen were trials or prospective cohorts, one was a retrospective cohort, and 4 were cross-sectional. Seven assessed QOL, 12 assessed depression, 8 assessed anxiety, and 1 assessed death by suicide. Three studies included trans-feminine people only; 7 included trans-masculine people only, and 10 included both. Three studies focused on adolescents. Hormone therapy was associated with increased QOL, decreased depression, and decreased anxiety. Associations were similar across gender identity and age. Certainty in this conclusion is limited by high risk of bias in study designs, small sample sizes, and confounding with other interventions. We could not draw any conclusions about death by suicide. Future studies should investigate the psychological benefits of hormone therapy among larger and more diverse groups of transgender people using study designs that more effectively isolate the effects of hormone treatment.

Key Words: Transgender, hormone therapy, sex hormones, mental health, systematic review

ISSN 2472-1972

Transgender people are those whose gender identity is different from the sex they were assigned at birth. Estimates of the size of the transgender population vary depending on how the data are collected [1]. In studies that rely on clinical records, estimates range between 1 and 30 people per 100 000 (0.001% to 0.03%) [2]. Studies that focus instead on self-report among nonclinical populations find estimates that range between 0.1% and 2% [2].

Many transgender people seek medical services to affirm their gender identity. According to the *Standards of Care for Transsexual, Transgender, and Gender Non-Conforming People* maintained by the World Professional Association for Transgender Health (WPATH), gender-affirming medical care is different for each individual and may include a variety of services and procedures, such as psychological support, hormone therapy, and surgeries [3]. Hormone therapy, which typically involves estrogens and anti-androgens for transgender women and other trans-feminine people and testosterone for transgender men and other trans-masculine people, is a common component of medical gender affirmation [4]. Because hormone treatment can have a powerful effect on physical appearance, it is often a priority for transgender people seeking medical gender affirmation [5]. Gender-affirming hormone therapy can be managed for most patients by primary care providers, as it typically involves long-term maintenance on doses similar to those used for cisgender patients with conditions such as hypogonadism [6, 7]. Some clinicians require a minimum period of psychological counseling before hormone therapy can be initiated, while others provide hormone therapy on the basis of informed consent [8].

The need for gender-affirming care is often characterized using psychiatric diagnoses such as gender dysphoria, which replaced gender identity disorder in the fifth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (DSM-5) [9]. The 11th International Classification of Diseases (ICD-11) replaces these terms with a diagnosis called gender incongruence (codes: HA60, HA61, HA6Z), which is located in a new chapter on sexual health. These changes clarify that the target of gender-affirming medical interventions is not the person's gender identity itself but rather the clinically significant distress that can accompany a misalignment between gender identity and sex assigned at birth [10]. Some countries have further underscored that transgender identity is not a pathology by recognizing gender affirmation as fundamental to the human right to self-definition and removing requirements that transgender people seeking gender-affirming medical care present with a diagnosis such as gender dysphoria [11].

Several previous reviews have indicated that gender-affirming hormone therapy is associated with psychological benefits that include reductions in depression and anxiety and improvements in quality of life (QOL) among transgender people [12-17]. Most of these reviews did not require a minimum duration of hormone therapy [14-17]. One review that did impose a minimum follow-up requirement is 10 years old [12]. The other that required a minimum of 3 months of therapy included only uncontrolled prospective cohorts, which resulted in a sample of only 3 studies [13]. A comprehensive review without a minimum follow-up period assessed gender-affirming hormone therapy and surgeries only in adolescents [17]. By requiring a minimum duration of hormone treatment but considering all ages and a variety of study designs, we sought to update and more completely summarize the growing evidence base regarding the relationship between gender-affirming hormone therapy and psychological outcomes in transgender people.

Search Strategy and Selection Criteria

This review is one of a series of systematic reviews on gender-affirming care conducted for WPATH to inform the eighth revision of the *Standards of Care*. The protocol is registered on PROSPERO (CRD42018115379) [18], and we followed the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) guidelines in reporting our findings [19].

We searched PubMed, Embase, and PyscINFO from inception to October 2018 and updated the search through June 10, 2020, for studies assessing QOL, depression, anxiety, and death by suicide among transgender participants of any age in the context of gender-affirming hormone therapy [20]. We also reviewed the reference lists of previous reviews and hand-searched the *International Journal of Transgenderism*. Using DistillerSR [21], 2 reviewers independently screened titles, abstracts, and full-text articles. Differences were resolved through consensus adjudication.

We included studies that evaluated the psychological effects of any testosterone, estrogen, or anti-androgen formulation used for gender affirmation. We also considered gonadotropin-releasing hormone (GnRH) analogues used as anti-androgens or for puberty delay. Study participants must have been on hormone therapy for at least 3 months in order to reflect a minimum time for expected onset of effects [3]. Health care provider supervision was not required. We excluded studies that did not state therapy type and duration, including the range for cross-sectional studies. We included studies regardless of language (the search terms were in English) and country of origin, and we accepted any study design except case reports.

We created standardized forms for data extraction using the Systematic Review Data Repository system. The data extracted included participant demographics; study design

and methods; hormone therapy type, dose, and duration; potential confounders such as gender-affirming surgery status; outcome scales [20]; and psychological outcomes. From studies that used the Short Form-36 Health Survey (SF-36) to measure QOL, we extracted scores in all domains [22]. For studies that used measures with depression or anxiety subscales, we extracted only the subscale scores corresponding to the psychological outcomes of interest (eg, the depression subscale of the Minnesota Multiphasic Personality Inventory [MMPI]). We extracted comparisons with cisgender controls or general population norms only when longitudinal findings in a transgender population or comparisons with an untreated transgender control group were not reported. We used WebPlotDigitizer to extract data reported only in figures [23].

Two reviewers independently assessed risk of bias [20]. For randomized controlled trials (RCTs), we used the revised Cochrane tool [24]. For non-randomized studies, we used the Cochrane Risk of Bias Assessment Tool for Non-Randomized Studies of Interventions (ROBINS-I) [25]. One reviewer graded strength of evidence for each outcome using the Agency for Healthcare Research and Quality Methods Guide for Conducting Comparative Effectiveness Reviews [26]. We considered the directionality and magnitude of effects reported in cross-sectional studies as additional context for our evaluation of evidence from trials and prospective and retrospective cohorts. Each strength of evidence assessment was confirmed by a second reviewer.

WPATH provided the research question and reviewed the protocol, evidence tables, and report. WPATH had no role in study design, data collection, analysis, interpretation, or drafting. The corresponding author had full access to all the data and had final responsibility for the decision to submit for publication. The authors are responsible for all content, and statements in this report do not necessarily reflect the official views of or imply endorsement by WPATH.

Results

We retrieved 1753 nonduplicate studies for the broader systematic review project of which this review was a part (Fig. 1). After screening and full-text review for the specific research question on the psychological effects of gender-affirming hormone therapy, 20 studies reported in 22 publications were included (Table 1): 1 RCT [27], 2 before-after trials [28, 29], 12 prospective cohorts reported in 13 publications [30-42], 1 retrospective cohort reported in 2 publications [43, 44], and 4 cross-sectional studies [45-48]. De Vries (2014) [35] reported on a subset of the participants in de Vries (2011) [34] who continued in care. We counted these publications as a single study but extracted and reported data separately because the characteristics of the study's adolescent population changed substantially in the period between the 2 publications. Similarly, Asscheman (2011) [44] reported on an extension of Asscheman (1989) [43]; we counted these as a single study but extracted data separately. In Table 1 and in the subsequent tables for each outcome, studies are ordered first by study design (RCTs, before-after trials, prospective cohorts, retrospective cohorts, and cross-sectional studies); within these categories, studies are presented in the following order according to how the study results were reported: adult transgender women only, adult transgender men only, adult transgender women and transgender men together, and transgender adolescents (no study reported separate results by gender identity for transgender youth). Where multiple studies shared the same study design and population, they are additionally ordered chronologically.

The time frame covered in the included studies began in 1972 [43], but most studies dated from post-2000. Eight studies were conducted in Italy [27-29, 31, 32, 36, 39, 41]; 2 each in Belgium [37, 48], the Netherlands [34, 35, 43, 44], the United States [30, 47], and Spain [38, 45]; and 1 in the United Kingdom [33], Turkey [42], and France [46]. One study recruited participants from Switzerland and Germany [40]. One study was part of the European Network for the Investigation of Gender Incongruence (ENIGI), which is a research collaborative between clinics providing gender-affirming care to transgender people in Ghent (Belgium), Amsterdam (Netherlands), Oslo (Norway), and Hamburg (Germany). The ENIGI study included in this review drew participants only from the Ghent clinic [37].

The study sizes ranged from 20 to 1331, although most had fewer than 60 participants. Fourteen studies reported on testosterone formulations in adult transgender men [27, 29, 31-33, 36, 39-46, 48]. These formulations were typically injectable testosterone cypionate or enanthate, although some studies used long-acting injectable testosterone undecanoate or daily transdermal gels. Ten studies reported on estrogen formulations in adult transgender women, usually in conjunction with an anti-androgen such as cyproterone acetate or spironolactone [28, 31, 33, 36, 37, 39, 43-47]. Estrogen formulations included transdermal, oral, or injectable estradiol (commonly estradiol valerate) or conjugated estrogens. Three studies reported on the psychological effects of GnRH therapy for puberty delay among mixed-gender groups of transgender adolescents [30, 34, 35, 38]. No study reported on hormone therapy among nonbinary people.

All studies that reported information about recruitment drew their participants largely or exclusively from specialized clinics dedicated to providing gender-affirming care for transgender people. These clinics were typically part of larger systems such as university hospitals. Clinic-based

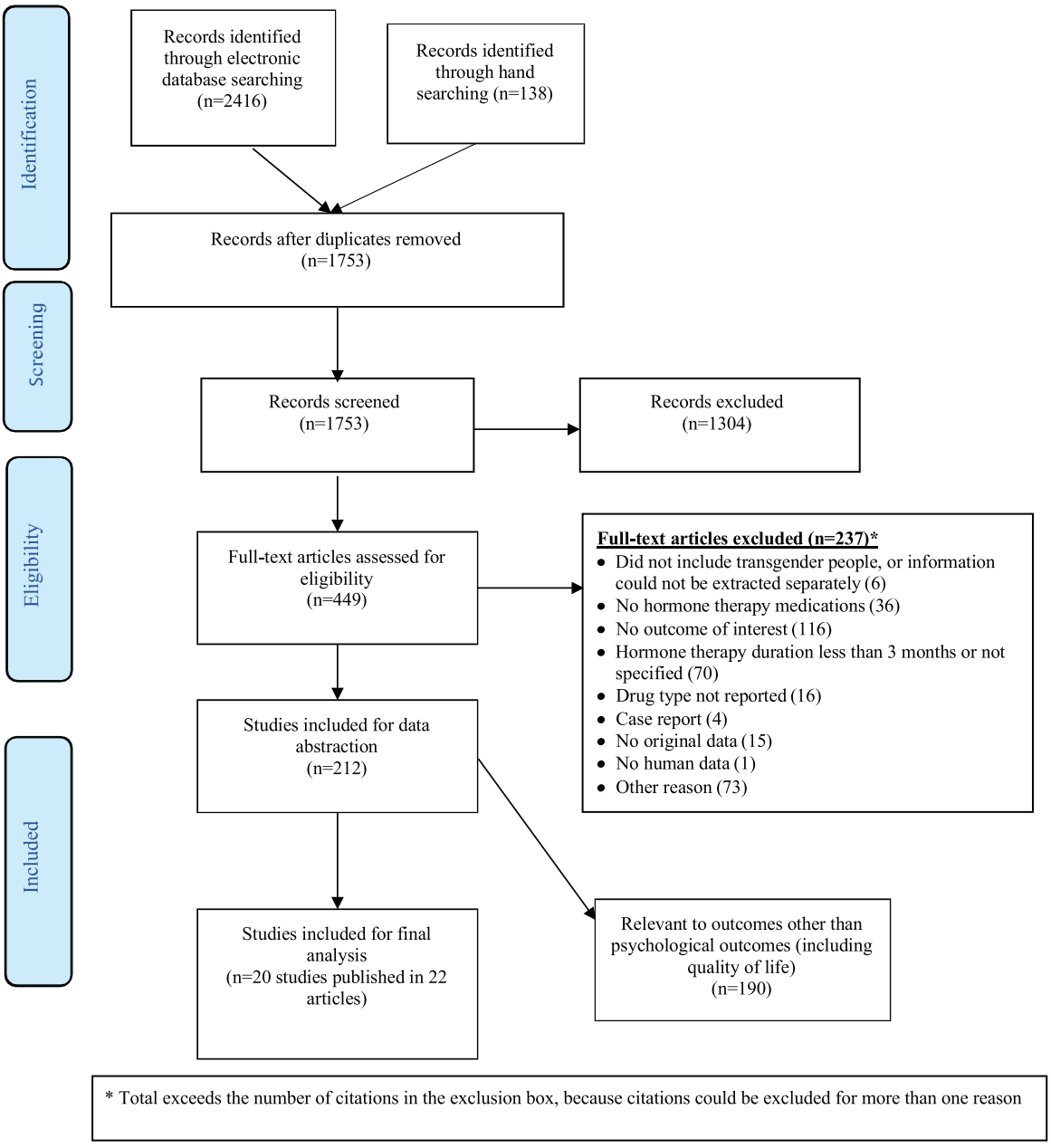


Figure 1. PRISMA flow diagram.

studies often applied strict eligibility criteria that included a period of psychiatric evaluation and a formal diagnosis of gender dysphoria before hormone therapy was initiated. Some studies also reported that psychological counseling was either available or required during the course of hormone therapy. In many cases, hormone therapy was considered a prerequisite for gender-affirming surgeries. The type and timing of gender-affirming surgeries and the proportion of participants for whom hormone therapy and surgeries were assessed simultaneously varied widely: some studies assessed only participants who had not had any type of gender-affirming surgery [27, 28, 30-32, 34, 36, 38-40, 42, 46, 47], while in others some or all participants underwent gender-affirming surgeries during the study period [29, 33, 35, 43-45, 48].

Quality of Life

Seven studies, including 1 RCT [27], 2 before-after trials [28, 29], 2 prospective cohorts [30, 39], and 2 cross-sectional studies [46, 48], assessed QOL (Table 2). An RCT found an improvement of approximately 5.5 points on a 10-point measure of life satisfaction across 3 groups of transgender men (n = 15 each) after 1 year of testosterone treatment ($P < 0.05$) [27]. A before-after trial similarly reported that life satisfaction scores almost

Table 1. Studies Reporting Effects of Gender-Affirming Hormone Therapy on Psychological Outcomes Among Transgender People

Author, year Location Study name	Study design	Start year	Transgender population	Overall N	Age in years	Baseline HT status	Outcomes	GAS status	Risk of bias
Pelusi, 2014 [27] Italy	Randomized controlled trial[a]	NR	Men	45	Mean: 29.5	No previous HT	QOL	No GAS before or during study	High
Gava, 2016 [28] Italy	Before-after trial	NR	Women	40	Mean: 3.2 (range, 19–55)	No previous HT	QOL, Depression	No GAS before or during study	Low
Gava, 2018 [29] Italy	Before-after trial[a]	NR	Men	50	Mean: 30.1 (range, 21–42)	No previous HT	QOL	72% (n = 36) had gonadectomy during study	Serious
Fuss, 2015 [37] Belgium ENIGI (NCT01072825)	Prospective cohort	2010	Women	20	Mean: 33.9 (range, 17–48)	No previous HT	Anxiety	NR	Serious
Costantino, 2013 [32] Italy	Prospective cohort	2001	Men	50	Mean: 29.8	No previous HT	Depression	No GAS before or during study	Serious
Motta, 2018 [41] Italy	Prospective cohort	2013	Men	52	Mean: 28.3	No previous HT	Anxiety	NR	Moderate
Turan, 2018 [42] Turkey	Prospective cohort[b]	NR	Men	37	Mean: 24.6	No previous HT	Depression, Anxiety	No GAS before or during study	Moderate
Metzger, 2019 [40] Switzerland, Germany	Prospective cohort[b]	2013	Men	23	Mean: 27.2 (range, 18–51)	No previous HT	Depression	No GAS before or during study	Moderate
Colizzi, 2014 [31] Italy	Prospective cohort	2008	Women and men	107	Mean: 29.2	No previous HT	Depression, Anxiety	No GAS before or during study	Low
Manieri, 2014 [39] Italy	Prospective cohort	NR	Women and men	83	Mean: 32.7 (women), 30.2 (men)	No previous HT	QOL	No GAS before or during study	Moderate
Fisher, 2016 [36] Italy	Prospective cohort	2012	Women and men	54	Mean: 32.5 (women), 26.3 (men)	No previous HT	Depression	No GAS before or during study	Low
Defreyne, 2018 [33] UK	Prospective cohort	2012	Women and men	155	Median: 27 (range, 18–52)	No previous HT	Depression, Anxiety	Some had GAS during study; % and type NR	Serious
Asscheman, 1989 [43] Netherlands	Retrospective cohort[b,d]	1972	Women and men	425	Median: 32 (women, range, 16–67); 25.4 (men, range, 16–54)	Previous HT for at least 6 months	Death by suicide	78% (n = 235) of transgender women had GAS during study; data NR for transgender men	Serious

Table 1. Continued

Author, year Location Study name	Study design	Start year	Transgender population	Overall N	Age in years	Baseline HT status	Outcomes	GAS status	Risk of bias
Asscheman, 2011 [44] Netherlands	Retrospective cohort[b,d]	1975	Women and men	1331	Mean: 31.4 (women, range, 16–76); 26.1 (men, range, 16–57)	Previous HT for at least 1 year	Death by suicide	87% (n = 834) of transgender women and 94% (n = 343) of transgender men had GAS during study	Serious
Leavitt, 1980 [47] US	Cross-sectional	1976	Women	41	Range, 18–35	54% (n = 22) on HT	Depression	No previous GAS	Serious
Wierckx, 2011 [48] Belgium	Cross-sectional[b]	2009	Men	47	Mean: 37 (range, 22–54)	100% on HT	QOL	100% had GAS, but not within previous year	Serious
Gómez-Gil, 2012 [45] Spain	Cross-sectional	NR	Women and men	187	Mean: 29.9 (range, 15–61)	64% (n = 120) on HT	Depression, Anxiety	42% (n = 79) of all participants and 64% (n = 77) of participants on HT had previous GAS	Serious
Gorin-Lazard, 2012 [46] France	Cross-sectional[b]	NR	Women and men	61	Mean: 34.7	72% (n = 44) on HT	QOL	No previous GAS	Serious
de Vries, 2011 [34] Netherlands	Prospective cohort	2000	Girls and boys	70	Mean: 14.8 (range, 11.3–18.6)	No previous HT	Depression, Anxiety	No GAS before or during study	Moderate
de Vries, 2014 [35] Netherlands	Prospective cohort[b,c]	2000	Girls and boys	55	Mean: 14.8 (range, 11.5–18.5)	No previous HT	Depression, Anxiety	100% had GAS during study	Serious
Achille, 2020 [30] US	Prospective cohort	2013	Girls and boys	50	Mean: 16.2	No previous HT	QOL, Depression	No GAS before or during study	Moderate
López de Lara, 2020 [38] Spain	Prospective cohort[b]	2018	Girls and boys	23	Mean: 16 (range, 14-18)	No previous HT	Depression, Anxiety	No GAS before or during study	Moderate

Abbreviations: ENIGI, European Network for the Investigation of Gender Incongruence; GAS, gender-affirming surgery; HT, hormone therapy; NR, not reported; QOL, quality of life.
[a]25 participants were included in both Pelusi [27] and Gava (2018) [29]
[b]Included a cisgender control group or a comparison to general population norms
[c]All participants were also included in de Vries (2011) [34]
[d]An unknown number of participants were included in both Asscheman (1989) [43] and Asscheman (2011) [44]

Table 2. Effects of Gender-Affirming Hormone Therapy on Quality of Life Among Transgender People

Author, year Study design	Transgender population	Treatment / comparison (n)	QOL measures	Length of treatment	Findings
Pelusi, 2014 [27] RCT[a]	Men	Testoviron depot (15) vs testosterone gel (15) vs testosterone undecanoate (15)	VAS (general life satisfaction)	54 weeks	Mean QOL scores increased from 2.8 to 8.5 ($P < 0.05$) in the testoviron depot arm, from 3.2 to 8.9 ($P < 0.05$) in the testosterone gel arm, and from 2.6 to 8.0 ($P < 0.05$) in the testosterone undecanoate arm.[d] There was no difference across arms.
Gava, 2016 [28] Before-after trial	Women	Cyproterone acetate + estradiol (20) vs leuprolide acetate + estradiol (20)	VAS (general life satisfaction) SF-36	12 months	Mean QOL scores did not change in either arm. No comparisons across arms were reported.
Gava, 2018 [29] Before-after trial[a]	Men	Testosterone undecanoate (25)[c] vs testosterone enanthate (25)[c]	VAS (general satisfaction)	5 years	Mean QOL scores increased from 4.3 ± 3.1 to 8.1 ± 1.8 ($P < 0.001$) in the testosterone undecanoate arm and from 4.3 ± 3.8 to 8.3 ± 1.7 ($P < 0.001$) in the testosterone enanthate arm. No comparisons across arms were reported.
Manieri, 2014 [39] Prospective cohort	Women	HT (56)	WHOQOL	12 months	Mean QOL scores increased from 62.5 to 72.2 ($P < 0.05$).[d]
Manieri, 2014 [39] Prospective cohort	Men	HT (27)	WHOQOL	12 months	Mean QOL scores did not change.
Wierckx, 2011 [48] Cross-sectional[b]	Men	HT (47)[c]	SF-36	At least 3 years	Mean QOL scores on the VT and MH subscales were lower for transgender men than cisgender men (VT subscale: 62.1 ± 20.7 vs 71.9 ± 18.3, $P = 0.002$; MH subscale: 72.6 ± 19.2 vs 79.3 ± 16.4, $P = 0.020$). There were no other differences between transgender men and either cisgender men or cisgender women.
Gorin-Lazard, 2012 [46] Cross-sectional[b]	Women and men	HT (44) vs no HT (17)	SF-36	Median: 20 months (range, 12–42 months)	Mean QOL scores were generally higher in the group receiving HT vs the group not receiving HT (MCS: 51.0 ± 7.7 vs 39.8 ± 12.7, $P = 0.003$; MH subscale: 76.4 ± 14.1 vs 59.1 ± 19.6, $P = 0.004$; RE subscale: 88.6 ± 22.7 vs 54.9 ± 40.7, $P = 0.001$; SF subscale: 83.2 ± 23.3 vs 69.9 ± 24.2, $P = 0.026$). There were no differences in the other subscales.
Achille, 2020 [30] Prospective cohort	Girls and boys	GnRH treatment + HT (47)	Q-LES-Q-SF	12 months	Mean QOL scores did not change.

Abbreviations: GnRH, gonadotropin-releasing hormone; HT, hormone therapy; MCS, Mental Component Summary; MH, mental health; QOL, quality of life; RCT, randomized controlled trial; RE, role functioning/emotional; SF, social functioning; SF-36, Short Form-36 Health Survey; VAS, visual analog scale; VT, vitality; WHOQOL, World Health Organization Quality of Life measure.

[a]10 participants on testosterone enanthate and 15 participants on testosterone undecanoate were included in both Pelusi [27] and Gava (2018) [29]

[b]Included a cisgender control group or a comparison to general population norms

[c]Included participants who had undergone gender-affirming surgery/surgeries, or surgery status not reported

[d]No standard deviations reported

doubled among transgender men (n = 50) over 5 years [29]. A prospective study found a 16% improvement in QOL scores among transgender women (n = 56) after 1 year of treatment ($P < 0.05$) but no change among transgender men (n = 27) [39]. Another before-after trial reported no difference in SF-36 scores among 2 groups of transgender women (n = 20 each) after 1 year [28]. Among adolescents, a mixed-gender prospective cohort (n = 50) showed no difference in QOL scores after a year of endocrine interventions, which included combinations of GnRH analogues and estrogen or testosterone formulations [30]. No study found that hormone therapy decreased QOL scores. We conclude that hormone therapy may improve QOL among transgender people. The strength of evidence for this conclusion is low due to concerns about bias in study designs, imprecision in measurement because of small sample sizes, and confounding by factors such as gender-affirming surgery status.

Depression

Twelve studies, including 1 before-after trial [28], 9 prospective cohorts [30-36, 38, 40, 42], and 2 cross-sectional studies [45, 47], assessed depression (Table 3). A prospective study found that the proportion of transgender men and transgender women (n = 107) showing symptoms of depression decreased from 42% to 22% over 12 months of treatment ($P < 0.001$) [31]. In 2 other prospective cohorts, Beck Depression Inventory (BDI-II) scores improved by more than half among both transgender men (n = 26) and transgender women (n = 28) after 24 months of therapy ($P < 0.001$) [36] and improved from 15.7 ± 12.3 to 8.1 ± 6.2 among transgender men (n = 23) after 6 months ($P < 0.001$) [40]. A fourth prospective study reported improvements of 1.05 points (95% CI: −1.87, −0.22) and 1.42 points (95% CI: −2.61, −0.24) on the 21-point Hospital Anxiety and Depression Scale (HADS) among 91 transgender women and 64 transgender men after 12 months ($P = 0.013$ and $P = 0.019$, respectively) [33]. A before-after trial, however, found no change in BDI-II scores among 2 groups of transgender women (n = 20 each) after 1 year [28]. Two prospective studies reported no difference among transgender men (n = 37) after 24 weeks [42] or among transgender men (n = 50) after 12 months [32], although in the latter study this outcome did not change from a baseline median of 0.0 ("not at all depressed") on an unvalidated 4-point scale. Among adolescents, 2 mixed-gender prospective cohorts (n = 50 and n = 23, respectively) showed improvements in depression scores after 1 year of treatment with GnRH analogues and estrogen or testosterone formulations (both $P < 0.001$) [30, 38]. Another prospective study reported that BDI scores improved almost by half among adolescents (n = 41) after a mean of 1.88 years of treatment with GnRH analogues to delay puberty ($P = 0.004$) [34]. The overall improvement after several subsequent years of testosterone or estrogen therapy in this cohort (n = 32) was smaller, however, resulting in no significant change from baseline [35]. No study found that hormone therapy increased depression. We conclude that hormone therapy may decrease depression among transgender people. The strength of evidence for this conclusion is low due to concerns about study designs, small sample sizes, and confounding.

Anxiety

Eight studies, including 7 prospective cohorts [31, 33-35, 37, 38, 41, 42] and 1 cross-sectional study [45], assessed anxiety (Table 4). One prospective study found that Symptom Checklist 90-Revised scores indicating a probable anxiety disorder among a mixed-gender group of adults (n = 107) improved from borderline to normal over 12 months ($P < 0.001$) [31]. Another prospective study, however, did not find a difference in HADS anxiety scores among either transgender men (n = 64) or transgender women (n = 91) after 1 year [33], and a third study reported no change in the number of transgender men (6/52, 12%) with a diagnosed anxiety disorder after 7 months [41]. Likewise, 2 other prospective studies found no difference in anxiety scores among transgender men (n = 37) after 24 weeks of treatment [42] or transgender women (n = 20) after 12 months [37], although this latter finding represented no change from a baseline median score of 0 (answering "no" to the question, "do you feel anxious?") on an unvalidated 3-point scale. Among adolescents, 1 prospective study saw mean anxiety scores in a mixed-gender group (n = 23) improve from 33.0 ± 7.2 to 18.5 ± 8.4 after 1 year ($P < 0.001$) [38], but another reported no changes in anxiety after approximately 2 years of puberty delay treatment with GnRH analogues and 4 years of hormone therapy (n = 32) [35]. No study found that hormone therapy increased anxiety. We conclude that hormone therapy may decrease anxiety among transgender people. The strength of evidence for this conclusion is low due to concerns about study designs, small sample sizes, and confounding.

Death by Suicide

One retrospective study reported in 2 publications assessed death by suicide (Table 5) [43, 44]. The first publication reported that 3 transgender women in the Amsterdam gender dysphoria study cohort (n = 303) died by suicide between 1972 and 1986 [43]. The authors calculated the number of suicide deaths expected in an age-matched stratum of

Table 3. Effects of Gender-Affirming Hormone Therapy on Depression Among Transgender People

Author, year Study design	Transgender population	Treatment / comparison (n)	Depression measures	Length of treatment	Findings
Gava, 2016 [28] Before-after trial	Women	Cyproterone acetate + estradiol (20) vs Leuprolide acetate + estradiol (20)	BDI-II	12 months	Mean depression scores did not change in either arm. No comparisons across arms were reported.
Fisher, 2016 [37] Prospective cohort	Women	HT (28)	BDI-II	24 months	Mean depression score decreased from 10.12 to 4.58 ($P < 0.001$).[d,e]
Defreyne, 2018 [33] Prospective cohort	Women	HT (91)[c]	HADS (depression subscale)	1 year	Median depression score decreased by 1.05 (95% CI: −1.87, −0.22) on a 21-point scale ($P = 0.013$).
Costantino, 2013 [32] Prospective cohort	Men	HT (50)	Ad hoc questionnaire	12 months	Depression score did not change from a median of 0.0 at baseline (IQR: 0.0, 1.0).
Fisher, 2016 [36] Prospective cohort	Men	HT (26)	BDI-II	24 months	Mean depression score decreased from 9.31 to 4.25 ($P < 0.001$).[d,e]
Defreyne, 2018 [33] Prospective cohort	Men	HT (64)[c]	HADS (depression subscale)	1 year	Median depression score decreased by 1.42 (95% CI: −2.61, −0.24) on a 21-point scale ($P = 0.019$).
Turan, 2018 [42] Prospective cohort[b]	Men	HT (37)	SCL-90-R (depression subscale)	24 weeks	Mean depression score did not change.
Metzger, 2019 [40] Prospective cohort[b]	Men	HT (23)	BDI-II	6 months	Mean depression score decreased from 15.7 ± 12.3 to 8.1 ± 6.2 ($P < 0.001$).
Colizzi, 2014 [31] Prospective cohort	Women and men	HT (107)	Zung SDS SCL-90-R (depression subscale)	12 months	Mean Zung SDS score improved from 48.40 ± 10.5 to 39.98 ± 10.79 ($P < 0.001$), and the proportion with Zung SDS scores indicating mild, moderate, or severe depression (vs no depression) decreased from 42% to 22% ($\chi^2 = 19.05$, $P < 0.001$). Mean SCL-90-R score decreased from 0.83 ± 0.74 to 0.51 ± 0.49 ($P < 0.001$), which represents an improvement from possible borderline depression to no depression.
Leavitt, 1980 [47] Cross-sectional	Women	HT (22) vs No HT (19)	MMPI (depression subscale)	At least 12 months	Mean depression score was lower in the group receiving HT vs the group not receiving HT (53.1 ± 14.7 vs 65.7 ± 11.2, $P = 0.004$).

Table 3. Continued

Author, year Study design	Transgender population	Treatment / comparison (n)	Depression measures	Length of treatment	Findings
Gómez-Gil, 2012 [45] Cross-sectional	Women and men	HT (120)[c] vs No HT (67)[c]	HADS (depression subscale)	Mean: 11.0 years (women, range, 1–46 years); 4.7 years (men, range, 1–22 years)	Mean depression score was lower in the group receiving HT vs the group not receiving HT (3.3 ± 3.2 vs 5.2 ± 4.2, $P = 0.002$).[f] The proportion with scores indicating depression (vs no depression) was larger in the group not receiving HT (31% vs 8%, $\chi^2 = 16.46$, $P = 0.001$).[f]
de Vries, 2011 [34] Prospective cohort	Girls and boys	GnRH treatment (41)	BDI	1.88 years	Mean depression score decreased from 8.31 ± 7.12 to 4.95 ± 6.72 ($P = 0.004$).
de Vries, 2014 [35] Prospective cohort[a,b]	Girls and boys	GnRH treatment + HT (32)[c]	BDI	5.9 years	Mean depression score did not change.
Achille, 2020 [30] Prospective cohort	Girls and boys	GnRH treatment + HT (47)	CESD-R, PHQ-9 (modified for adolescents)	12 months	Mean CESD-R score decreased from 21.4 to 13.9 ($P < 0.001$);[d] a score of <16 indicates no clinical depression. Mean PHQ-9 score decreased from 9.0 to 5.4 ($P < 0.001$).[d]
López de Lara, 2020 [38] Prospective cohort[b]	Girls and boys	GnRH treatment + HT (23)	BDI-II	1 year	Mean depression score decreased from 19.3 ± 5.5 to 9.7 ± 3.9 ($P < 0.001$).

Abbreviations: BDI/BDI-II, Beck Depression Inventory; GAS, gender-affirming surgery; GnRH, gonadotropin-releasing hormone; HADS, Hospital Anxiety and Depression Scale; HT, hormone therapy; IQR, interquartile range; MMPI, Minnesota Multiphasic Personality Inventory; NA, not applicable; SCL-90-R, Symptom Checklist 90-Revised; Zung SDS, Zung Self-Rating Depression Scale.

[a]All participants were also included in de Vries (2011) [34]

[b]Included a cisgender control group or a comparison to general population norms

[c]Included participants who had undergone gender-affirming surgery/surgeries, or surgery status not reported

[d]No standard deviations reported

[e]Adjusted for age, gender role, and surgery status

[f]Adjusted for age, gender, and education level

Table 4. Effects of Gender-Affirming Hormone Therapy on Anxiety Among Transgender People

Author, year	Transgender population	Treatment / comparison (n)	Anxiety measures	Length of treatment	Findings
Fuss, 2015 [37] Prospective cohort	Women	HT (20)[c]	Ad hoc questionnaire	12 months	Anxiety score did not change from a median of 0.0 at baseline.
Defreyne, 2018 [33] Prospective cohort	Women	HT (91)[c]	HADS (anxiety subscale)	1 year	Median anxiety score did not change.
Defreyne, 2018 [33] Prospective cohort	Men	HT (64)[c]	HADS (anxiety subscale)	1 year	Median anxiety score did not change.
Motta, 2018 [41] Prospective cohort	Men	HT (46)[c]	DSM	7 months	Proportion diagnosed with an anxiety disorder (6/46, 12%) did not change.
Turan, 2018 [42] Prospective cohort[b]	Men	HT (37)	SCL-90-R (anxiety subscale)	24 weeks	Mean anxiety score did not change.
Colizzi, 2014 [31] Prospective cohort	Women and men	HT (107)	SCL-90-R (anxiety subscale) Zung SAS	12 months	Mean SCL-90-R score decreased from 1.05 ± 0.95 to 0.54 ± 0.56 ($P < 0.001$), which represents an improvement from borderline anxiety disorder to no anxiety disorder. Mean Zung SAS score improved from 44.91 ± 9.59 to 37.90 ± 8.97 ($P < 0.001$), and the proportion with Zung SAS scores indicating mild, moderate, or severe anxiety (vs no anxiety) decreased from 50% to 17% ($\chi^2 = 33.03$, $P < 0.001$).
Gómez-Gil, 2012 [45] Cross-sectional	Women and men	HT (120)[c] vs No HT (67)[c]	HADS (anxiety subscale) SADS	Mean: 11.0 years (women, range, 1-46 years); 4.7 years (men, range, 1-22 years)	Mean HADS and SADS scores were lower in the group receiving HT vs the group not receiving HT (6.4 ± 3.7 vs 9.0 ± 4.0, $P = 0.001$; 8.5 ± 7.8 vs 11.0 ± 7.3, $P = 0.038$, respectively).[d] The proportion with scores indicating anxiety (vs no anxiety) was higher in the group not receiving HT ($\chi^2 = 14.46$, $P < 0.001$).[d]
de Vries, 2011 [34] Prospective cohort	Girls and boys	GnRH treatment (41)	STAI (trait subscale)	1.88 years	Mean anxiety score did not change.
de Vries, 2014 [35] Prospective cohort[a,b]	Girls and boys	GnRH treatment + HT (32)[c]	STAI (trait subscale)	5.9 years	Mean anxiety score did not change.
López de Lara, 2020 [38] Prospective cohort[b]	Girls and boys	GnRH treatment + HT (23)	STAI (trait subscale)	1 year	Mean anxiety score decreased from 33.0 ± 7.2 to 18.5 ± 8.4 ($P < 0.001$).

Abbreviations: BAI, Beck Anxiety Inventory; DSM, Diagnostic and Statistical Manual of Mental Disorders; GAS, gender-affirming surgery; GnRH, gonadotropin-releasing hormone; HADS, Hospital Anxiety and Depression Scale; HT, hormone therapy; IQR, interquartile range; SADS, Social Avoidance and Distress Scale; SCL-90-R, Symptom Checklist 90-Revised; STAI, State-Trait Anxiety Inventory; Zung SAS, Zung Self-Rating Anxiety Scale.

[a]All participants were also included in de Vries (2011) [34]

[b]Included a cisgender control group or a comparison to general population norms

[c]Included participants who have undergone gender-affirming surgery/surgeries, or surgery status not reported

[d]Adjusted for age, gender, and education level

the general male Dutch population over this period to be 0.208. No data were reported for transgender men (n = 122). An update to this study reported 17 deaths by suicide among transgender women (n = 966) and 1 among transgender men (n = 365) between 1975 and 2007 [44]. The age- and sex-stratified standardized mortality ratios were 5.70 (95% CI: 4.93, 6.54) and 2.22 (95% CI: 0.53, 6.18), respectively. The risk of bias for this study was serious due to the difficulty of identifying appropriate comparison groups and uncontrolled confounding by surgery

Table 5. Effects of Gender-Affirming Hormone Therapy on Death by Suicide Among Transgender People

Author, year	Transgender population	Treatment / comparison (n)	Measures	Length of treatment	Findings
Asscheman, 1989 [43] Retrospective cohort[a,b]	Women	HT (303)[c]	Death by suicide (confirmed by autopsy report)	Median: 4.4 years (range, 6 months to 13 years)	3 transgender women (1%) died by suicide between 1972 and 1986. The adjusted number of suicide deaths expected among the general Dutch male population was 0.208.
Asscheman, 2011 [44] Retrospective cohort[a,b]	Women	HT (966)[c]	Death by suicide (confirmed by medical report or physician information)	Median: 18.6 years (range, 0.7–44.5 years)	17 transgender women (2%) died by suicide between 1975 and 2007. The age-stratified SMR compared to the general Dutch male population was 5.70 (95% CI: 4.93, 6.54).
Asscheman, 1989 [43] Retrospective cohort[a,b]	Men	HT (122)[c]	Death by suicide (confirmation procedure NR)	Median: 3.6 years (range, 6 months to 13 years)	No deaths by suicide among transgender men were reported during the study period.
Asscheman, 2011 [44] Retrospective cohort[a,b]	Men	HT (365)[c]	Death by suicide (confirmed by medical report or physician information)	Median: 18.4 years (range, 4.7–42.6 years)	1 transgender man (0.3%) died by suicide between 1975 and 2007. The age-stratified SMR compared to the general Dutch female population was 2.22 (95% CI: 0.53, 6.18).

Abbreviations: HT, hormone therapy; NR, not reported; SMR, standardized mortality ratio.
[a]An unknown number of participants were included in both Asscheman (1989) [43] and Asscheman (2011) [44]
[b]Included a cisgender control group or a comparison to general population norms
[c]Includes participants who had undergone gender-affirming surgery/surgeries, or surgery status not reported

status and socioeconomic variables such as unemployment. We cannot draw any conclusions on the basis of this single study about whether hormone therapy affects death by suicide among transgender people.

Discussion

This systematic review of 20 studies found evidence that gender-affirming hormone therapy may be associated with improvements in QOL scores and decreases in depression and anxiety symptoms among transgender people. Associations were similar across gender identity and age. The strength of evidence for these conclusions is low due to methodological limitations (Table 6). It was impossible to draw conclusions about the effects of hormone therapy on death by suicide.

Uncontrolled confounding was a major limitation in this literature. Many studies simultaneously assessed different types of gender-affirming care and did not control for gender-affirming surgery status, making it difficult to isolate the effects of hormone therapy. Others failed to report complete information about surgery status. Additional factors that may influence both access to care and psychological outcomes, including extent of social or legal gender affirmation and exposure to determinants of health such as discrimination, were typically not considered. In addition, some evidence indicates that cyproterone acetate, a common anti-androgen assessed in many studies alongside estrogen therapy, may increase depression, which may be a source of confounding [49].

Another source of potential bias was recruitment of participants from specialized clinics that impose strict diagnostic criteria as a prerequisite for gender-affirming care. The dual role of clinicians and researchers as both gatekeepers and investigators may force transgender study participants to over- or understate aspects of their mental health in order to access gender-affirming care [8]. Similarly, transgender clinic patients may feel that they cannot opt out of research-related activities, which is a serious concern for the validity of psychological outcome measurements.

Clinic-based recruitment also overlooks transgender people who cannot access these clinics for financial or other reasons and misses those whose need for gender affirmation does not fit into current medical models. This is a particular concern for nonbinary and other gender-diverse people, for whom a model of gender affirmation as a linear transition from one binary gender to another is inaccurate [50].

Most studies used well-known scales for measuring psychological outcomes. None of these scales, however, have been specifically validated for use in transgender populations [51]. Furthermore, many scales are normed

Table 6. Strength of Evidence of Studies that Evaluate the Psychological Effects of Hormone Therapy Among Transgender People

Outcome	Number of studies (n)	Strength of evidence	Summary[a]
Quality of life	1 randomized controlled trial [27] (45)[b] 2 before-after trials [28, 29] (65)[b] 2 prospective cohorts [30, 39] (133) 2 cross-sectional studies [46, 48] (108)	Low[e]	Hormone therapy may improve quality of life among transgender people.[g]
Depression	1 before-after trial [28] (40) 9 prospective cohorts [30-36, 38, 40, 42] (569)[c] 2 cross-sectional [45, 47] (228)	Low[e]	Hormone therapy may alleviate depression among transgender people.[g]
Anxiety	7 prospective cohorts [31, 33-35, 37, 38, 41, 42] (464)[c] 1 cross-sectional [45] (187)	Low[e]	Hormone therapy may alleviate anxiety among transgender people.[g]
Death by suicide	1 retrospective cohort [43, 44] (1756)[d]	Insufficient[f]	There is insufficient evidence to draw a conclusion about the effect of hormone therapy on death by suicide among transgender people.

[a]Due to similarity of findings, the summary is the same for transgender men and transgender women and for adolescents and adults
[b]25 participants are included in both Pelusi [27] and Gava (2018) [29] and are counted once
[c]All 55 participants in de Vries (2014) [35] were also included among the 70 participants in de Vries (2011) [34] and are counted once
[d]An unknown number of participants were included in both Asscheman (1989) [43] and Asscheman (2011), [44] so the unique sample size is smaller than indicated here
[e]Evidence downgraded due to study limitations, including uncontrolled confounding, and imprecision because of small sample sizes
[f]Evidence downgraded due to study limitations, including confounding and a lack of meaningful comparison groups, and imprecision in measurement of a rare event
[g]The body of cross-sectional evidence tended to align with the conclusion

separately for (presumed cisgender) men and women [52]. Inconsistency in identification of appropriate general population norms hinders comparisons between transgender and cisgender groups, which is a major related research question that requires further investigation.

Beyond methodological concerns in the studies we assessed, our review has other limitations. First, it is likely subject to publication bias, as we may have missed studies not published in the peer-reviewed literature. Second, a number of potentially relevant studies could not be included because the authors did not report on a minimum of 3 months of treatment or did not clearly state the type and/or duration of therapy, including the range for cross-sectional studies [53-65]. Finally, even where outcome measurements were similar across studies, heterogeneity in study designs, study populations, intervention characteristics, and reporting of results (ie, some studies reported results separately by gender identity, while others did not), prevented us from quantitatively pooling results.

More research is needed to further explore the relationship between gender-affirming hormone therapy and QOL, death by suicide, and other psychological outcomes, especially among adolescents. Future studies should investigate these outcomes in larger groups of diverse participants recruited outside clinical settings. Studies assessing the relationship between gender-affirming hormone therapy and mental health outcomes in transgender populations should be prospective or use strong quasi-experimental designs; consistently report type, dose, and duration of hormone therapy; adjust for possible confounding by gender-affirming surgery status; control for other variables that may independently influence psychological outcomes; and report results separately by gender identity. Despite the limitations of the available evidence, however, our review indicates that gender-affirming hormone therapy is likely associated with improvements in QOL, depression, and anxiety. No studies showed that hormone therapy harms mental health or quality of life among transgender people. These benefits make hormone therapy an essential component of care that promotes the health and well-being of transgender people.

Acknowledgments

Financial Support: This review was partly funded by the World Professional Association for Transgender Health.

Author Contributions: R.S. developed and implemented the search strategy with input from K.B., L.W., and K.R. K.B., L.W., R.S., V.D., K.M., and K.R. screened and assessed studies, extracted data, and graded strength of evidence. K.B. wrote the report, which was reviewed by all co-authors.

Additional Information

Correspondence: Kellan E. Baker, MPH, MA, Department of Health Policy and Management, Johns Hopkins Bloomberg School of Public Health, 624 North Broadway, Baltimore, MD 21205, USA. Email: kbaker39@jhu.edu.

Disclosure Summary: This review was partially funded by the World Professional Association for Transgender Health (WPATH). The authors of this manuscript are responsible for its content. Statements in the manuscript do not necessarily reflect the official views of or imply endorsement by WPATH.

Data Availability: Some or all data generated or analyzed during this study are included in this published article or in the data repository listed in the References.

References

1. Collin L, Reisner SL, Tangpricha V, Goodman M. Prevalence of transgender depends on the "case" definition: a systematic review. *J Sex Med*. 2016;**13**(4):613-626.
2. Goodman M, Adams N, Corneil T, Kreukels B, Motmans J, Coleman E. Size and distribution of transgender and gender nonconforming populations: a narrative review. *Endocrinol Metab Clin North Am*. 2019;**48**(2):303-321.
3. Coleman E, Bockting W, Botzer M, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgenderism*. 2012;**13**(4):165-232.
4. Hembree WC, Cohen-Kettenis PT, Gooren L, et al. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an Endocrine Society Clinical Practice Guideline. *J Clin Endocrinol Metab*. 2017;**102**(11):3869-3903.
5. James SE, Herman JL, Rankin S, Keisling M, Mottet L, Anafi M. *The Report of the 2015 U.S. Transgender Survey*. National Center for Transgender Equality; 2016.
6. Deutsch MB, ed. Guidelines for the primary and gender-affirming care of transgender and gender nonbinary people. 2016. Accessed December 19, 2020. https://transcare.ucsf.edu/guidelines
7. Wylie K, Knudson G, Khan SI, Bonierbale M, Watanyusakul S, Baral S. Serving transgender people: clinical care considerations and service delivery models in transgender health. *Lancet*. 2016;**388**(10042):401-411.
8. Schulz SL. The informed consent model of transgender care: an alternative to the diagnosis of gender dysphoria. *J Humanist Psychol*. 2018;**58**(1):72-92.
9. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. American Psychiatric Association; 2013.
10. Robles R, Fresán A, Vega-Ramírez H, et al. Removing transgender identity from the classification of mental disorders: a Mexican field study for ICD-11. *Lancet Psychiatry*. 2016;**3**(9):850-859.
11. Arístegui I, Radusky PD, Zalazar V, Romero M, Schwartz J, Sued O. Impact of the Gender Identity Law in Argentinean transgender women. *Int J Transgenderism*. 2017;**18**(4):446-456.
12. Murad MH, Elamin MB, Garcia MZ, et al. Hormonal therapy and sex reassignment: a systematic review and meta-analysis of quality of life and psychosocial outcomes. *Clin Endocrinol (Oxf)*. 2010;**72**(2):214-231.
13. White Hughto JM, Reisner SL. A systematic review of the effects of hormone therapy on psychological functioning and quality of life in transgender individuals. *Transgend Health*. 2016;**1**(1):21-31.
14. Costa R, Colizzi M. The effect of cross-sex hormonal treatment on gender dysphoria individuals' mental health: a systematic review. *Neuropsychiatr Dis Treat*. 2016;**12**:1953-1966.
15. Nobili A, Glazebrook C, Arcelus J. Quality of life of treatment-seeking transgender adults: a systematic review and meta-analysis. *Rev Endocr Metab Disord*. 2018;**19**(3):199-220.
16. Rowniak S, Bolt L, Sharifi C. Effect of cross-sex hormones on the quality of life, depression and anxiety of transgender individuals: a quantitative systematic review. *JBI Database System Rev Implement Rep*. 2019;**17**(9):1826-1854.
17. Mahfouda S, Moore JK, Siafarikas A, et al. Gender-affirming hormones and surgery in transgender children and adolescents. *Lancet Diabetes Endocrinol*. 2019;**7**(6):484-498.
18. Sharma R, Robinson K, Wilson L, Baker KE. Effects of hormone therapy in transgender people. Accessed December 19, 2020. https://www.crd.york.ac.uk/prospero/display_record.php?RecordID=115379
19. Moher D, Liberati A, Tetzlaff J, Altman DG; PRISMA Group. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *BMJ*. 2009;**339**:b2535.
20. Baker KE, Wilson LM, Sharma R, Dukhanin V, McArthur K, Robinson KA. Data associated with the publication: Hormone therapy, mental health, and quality of life among transgender people: a systematic review. *Johns Hopkins Univ Data Arch*. **V1**. doi: 10.7281/T1/E70MXR.
21. *Evidence Partners*. DistillerSR [software]; 2020.
22. Ware JE Jr, Kosinski M, Bayliss MS, McHorney CA, Rogers WH, Raczek A. Comparison of methods for the scoring and statistical analysis of SF-36 health profile and summary measures: summary of results from the Medical Outcomes Study. *Med Care*. 1995;**33**(4 Suppl):AS264-AS279.
23. Rohatgi A. WebPlotDigitizer: an HTML5-based online tool for to extract numerical data from plot images. 2020. https://automeris.io/WebPlotDigitizer/index.html
24. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. *BMJ*. 2019;**366**:l4898.
25. Sterne JA, Hernán MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*. 2016;**355**:i4919.
26. Berkman ND, Lohr KN, Ansari MT, et al. Grading the strength of a body of evidence when assessing health care interventions: an EPC update. *J Clin Epidemiol*. 2015;**68**(11):1312-1324.
27. Pelusi C, Costantino A, Martelli V, et al. Effects of three different testosterone formulations in female-to-male transsexual persons. *J Sex Med*. 2014;**11**(12):3002-3011.
28. Gava G, Cerpolini S, Martelli V, Battista G, Seracchioli R, Meriggiola MC. Cyproterone acetate vs leuprolide acetate in combination with transdermal oestradiol in transwomen: a comparison of safety and effectiveness. *Clin Endocrinol (Oxf)*. 2016;**85**(2):239-246.
29. Gava G, Mancini I, Cerpolini S, Baldassarre M, Seracchioli R, Meriggiola MC. Testosterone undecanoate and testosterone

enanthate injections are both effective and safe in transmen over 5 years of administration. *Clin Endocrinol (Oxf)*. 2018;**89**(6):878-886.
30. Achille C, Taggart T, Eaton NR, et al. Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: preliminary results. *Int J Pediatr Endocrinol*. 2020;**2020**:8.
31. Colizzi M, Costa R, Todarello O. Transsexual patients' psychiatric comorbidity and positive effect of cross-sex hormonal treatment on mental health: results from a longitudinal study. *Psychoneuroendocrinology*. 2014;**39**:65-73.
32. Costantino A, Cerpolini S, Alvisi S, Morselli PG, Venturoli S, Meriggiola MC. A prospective study on sexual function and mood in female-to-male transsexuals during testosterone administration and after sex reassignment surgery. *J Sex Marital Ther*. 2013;**39**(4):321-335.
33. Defreyne J, T'Sjoen G, Bouman WP, Brewin N, Arcelus J. Prospective evaluation of self-reported aggression in transgender persons. *J Sex Med*. 2018;**15**(5):768-776.
34. de Vries AL, Steensma TD, Doreleijers TA, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. *J Sex Med*. 2011;**8**(8):2276-2283.
35. de Vries AL, McGuire JK, Steensma TD, Wagenaar EC, Doreleijers TA, Cohen-Kettenis PT. Young adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*. 2014;**134**(4):696-704.
36. Fisher AD, Castellini G, Ristori J, et al. Cross-sex hormone treatment and psychobiological changes in transsexual persons: two-year follow-up data. *J Clin Endocrinol Metab*. 2016;**101**(11):4260-4269.
37. Fuss J, Hellweg R, Van Caenegem E, et al. Cross-sex hormone treatment in male-to-female transsexual persons reduces serum brain-derived neurotrophic factor (BDNF). *Eur Neuropsychopharmacol*. 2015;**25**(1):95-99.
38. López de Lara D, Pérez Rodríguez O, Cuellar Flores I, et al. Evaluación psicosocial en adolescentes transgénero. *An Pediatría*. 2020;**93**(1):41-48.
39. Manieri C, Castellano E, Crespi C, et al. Medical treatment of subjects with gender identity disorder: the experience in an Italian Public Health Center. *Int J Transgenderism*. 2014;**15**(2):53-65.
40. Metzger NY, Boettger S. The effect of testosterone therapy on personality traits of trans men: a controlled prospective study in Germany and Switzerland. *Psychiatry Res*. 2019;**276**:31-38.
41. Motta G, Crespi C, Mineccia V, Brustio PR, Manieri C, Lanfranco F. Does testosterone treatment increase anger expression in a population of transgender men? *J Sex Med*. 2018;**15**(1):94-101.
42. Turan Ş, Aksoy Poyraz C, Usta Sağlam NG, et al. Alterations in body uneasiness, eating attitudes, and psychopathology before and after cross-sex hormonal treatment in patients with female-to-male gender dysphoria. *Arch Sex Behav*. 2018;**47**(8):2349-2361.
43. Asscheman H, Gooren LJ, Eklund PL. Mortality and morbidity in transsexual patients with cross-gender hormone treatment. *Metabolism*. 1989;**38**(9):869-873.
44. Asscheman H, Giltay EJ, Megens JA, de Ronde WP, van Trotsenburg MA, Gooren LJ. A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones. *Eur J Endocrinol*. 2011;**164**(4):635-642.
45. Gómez-Gil E, Zubiaurre-Elorza L, Esteva I, et al. Hormone-treated transsexuals report less social distress, anxiety and depression. *Psychoneuroendocrinology*. 2012;**37**(5):662-670.
46. Gorin-Lazard A, Baumstarck K, Boyer L, et al. Is hormonal therapy associated with better quality of life in transsexuals? A cross-sectional study. *J Sex Med*. 2012;**9**(2):531-541.
47. Leavitt F, Berger JC, Hoeppner JA, Northrop G. Presurgical adjustment in male transsexuals with and without hormonal treatment. *J Nerv Ment Dis*. 1980;**168**(11):693-697.
48. Wierckx K, Van Caenegem E, Elaut E, et al. Quality of life and sexual health after sex reassignment surgery in transsexual men. *J Sex Med*. 2011;**8**(12):3379-3388.
49. Heinemann LA, Will-Shahab L, van Kesteren P, Gooren LJ; Collaborating Centers. Safety of cyproterone acetate: report of active surveillance. *Pharmacoepidemiol Drug Saf*. 1997;**6**(3):169-178.
50. Reisner SL, Hughto JMW. Comparing the health of non-binary and binary transgender adults in a statewide non-probability sample. *Plos One*. 2019;**14**(8):e0221583.
51. Thompson HM, Reisner SL, VanKim N, Raymond HF. Quality-of-life measurement: assessing the WHOQOL-BREF scale in a sample of high-HIV-risk transgender women in San Francisco, California. *Int J Transgend*. 2015;**16**(1):36-48.
52. Webb A, Heyne G, Holmes J, Peta J. Assessment norms for gender and implications for transgender, nonbinary populations. *Division 44 Newsletter*. 2016. Accessed June 9, 2020. https://www.apadivisions.org/division-44/publications/newsletters/division/2016/04/nonbinary-populations
53. Heylens G, Verroken C, De Cock S, T'Sjoen G, De Cuypere G. Effects of different steps in gender reassignment therapy on psychopathology: a prospective study of persons with a gender identity disorder. *J Sex Med*. 2014;**11**(1):119-126.
54. Gorin-Lazard A, Baumstarck K, Boyer L, et al. Hormonal therapy is associated with better self-esteem, mood, and quality of life in transsexuals. *J Nerv Ment Dis*. 2013;**201**(11):996-1000.
55. Gómez-Gil E, Vidal-Hagemeijer A, Salamero M. MMPI-2 characteristics of transsexuals requesting sex reassignment: comparison of patients in prehormonal and presurgical phases. *J Pers Assess*. 2008;**90**(4):368-374.
56. Oda H, Kinoshita T. Efficacy of hormonal and mental treatments with MMPI in FtM individuals: cross-sectional and longitudinal studies. *BMC Psychiatry*. 2017;**17**(1):256.
57. Elaut E, De Cuypere G, De Sutter P, et al. Hypoactive sexual desire in transsexual women: prevalence and association with testosterone levels. *Eur J Endocrinol*. 2008;**158**(3):393-399.
58. Warmuz-Stangierska I, Stangierski A, Ziemnicka K, et al. Emotional functions in transsexuals after the first step in physical transformation. *Endokrynol Pol*. 2015;**66**(1):47-52.
59. Colton Meier SL, Fitzgerald KM, Pardo ST, Babcock J. The effects of hormonal gender affirmation treatment on mental health in female-to-male transsexuals. *J Gay Lesbian Ment Health*. 2011;**15**(3):281-299.

60. Davis SA, Colton Meier S. Effects of testosterone treatment and chest reconstruction surgery on mental health and sexuality in female-to-male transgender people. *Int J Sex Health*. 2014;**26**(2):113-128.
61. Keo-Meier CL, Herman LI, Reisner SL, Pardo ST, Sharp C, Babcock JC. Testosterone treatment and MMPI-2 improvement in transgender men: a prospective controlled study. *J Consult Clin Psychol*. 2015;**83**(1):143-156.
62. Newfield E, Hart S, Dibble S, Kohler L. Female-to-male transgender quality of life. *Qual Life Res*. 2006;**15**(9):1447-1457.
63. Gooren LJ, Sungkaew T, Giltay EJ, Guadamuz TE. Cross-sex hormone use, functional health and mental well-being among transgender men (Toms) and Transgender Women (Kathoeys) in Thailand. *Cult Health Sex*. 2015;**17**(1):92-103.
64. van Kesteren PJ, Asscheman H, Megens JA, Gooren LJ. Mortality and morbidity in transsexual subjects treated with cross-sex hormones. *Clin Endocrinol (Oxf)*. 1997;**47**(3):337-342.
65. Wiepjes CM, den Heijer M, Bremmer MA, et al. Trends in suicide death risk in transgender people: results from the Amsterdam Cohort of Gender Dysphoria study (1972-2017). *Acta Psychiatr Scand*. 2020;**141**(6):486-491.

Debate: Different strokes for different folks

Article *in* Child and Adolescent Mental Health · May 2019

DOI: 10.1111/camh.12330

CITATIONS
6

READS
1,059

1 author:

Kenneth J. Zucker
University of Toronto
282 PUBLICATIONS **14,078** CITATIONS

SEE PROFILE

Child and Adolescent Mental Health

Child and Adolescent Mental Health Volume **, No. *, 2019, pp. **–** doi:10.1111/camh.12330

Debate: Different strokes for different folks

Kenneth J. Zucker

Department of Psychiatry, University of Toronto, Toronto, ON, Canada

A gender social transition in prepubertal children is a form of psychosocial treatment that aims to reduce gender dysphoria, but with the likely consequence of subsequent (lifelong) biomedical treatments as well (gender-affirming hormonal treatment and surgery). Gender social transition of prepubertal children will increase dramatically the rate of gender dysphoria persistence when compared to follow-up studies of children with gender dysphoria who did not receive this type of psychosocial intervention and, oddly enough, might be characterized as iatrogenic. Parents who bring their children for clinical care hold different philosophical views on what is the best way to help reduce the gender dysphoria, which require both respect and understanding.

Keywords: Gender identity; gender dysphoria; psychosocial treatment

The proverbial saying 'Different strokes for different folks' (The Oxford Dictionary of Phrase and Fable, 2006) reflects well the contemporary clinical debate on best-practice therapeutics for children with gender dysphoria. It reflects not only the variation in the philosophical and theoretical perspectives of front-line clinicians, but also variation in the philosophical belief systems of parents who bring their children to mental health professionals for clinical advice and care.

For prepubertal children with gender dysphoria, I would argue that there are three main approaches to therapeutics, which I list here in chronological/historical order: (a) active psychosocial treatment to reduce gender dysphoria so that the child's eventual gender identity is more congruent with her or his biological sex (thus obviating the necessity for what some now call 'gender-affirming' hormonal and surgical treatment); (b) 'wait-and-see' or 'watchful waiting', which makes the assumption that it is difficult to predict what the long-term outcome will be and so, well, the clinician should not recommend very much one way or the other; and (c) gender social transition, in which the child's 'social' gender identity is shifted from the gender assigned at birth to the putative desired gender (e.g., change in name, change in pronoun usage, and change in other phenotypic social attributes, such as hair-style and clothing-style that mark one's gender to significant others). Dreger (2009) characterized the first approach the 'therapeutic' model and the third approach the 'accommodation' model.

These rather marked variations in the type of psychosocial treatment considered to be in the best interest of the child reflect deep structure variations in theoretical perspectives on the nature and nurture of psychosexual differentiation (see the edited volume by Drescher & Byne, 2012). On the one hand, the first approach assumes that, for young children with gender dysphoria, gender identity is not fixed or 'locked in' at an early age and that there is a much greater degree of malleability and plasticity than might be the case for both adolescents and adults with gender dysphoria. On the other hand, the third approach assumes that gender identity is fixed and locked in at a very early age because of underlying biological mechanisms. One of the most well-known children with gender dysphoria, 'Jazz Jennings', has promulgated this view in her book, written for children, 'I Am Jazz' (Herthel & Jennings, 2014) where Jazz writes 'I have a girl brain but a boy body….I was born this way!'

As noted in several guideline reviews on clinical practice for the treatment of children with gender dysphoria (AACAP Practice Parameter on Gay, Lesbian, or Bisexual Sexual Orientation, Gender Nonconformity, and Gender Discordance in Children and Adolescents, 2012; American Psychological Association, 2015; Byne et al., 2012), the field suffers from a vexing problem: There are no randomized controlled trials (RCT) of different treatment approaches, so the front-line clinician has to rely on lower-order levels of evidence in deciding on what the optimal approach to treatment might be. One quote is sufficient to document this point: 'Different clinical approaches have been advocated for childhood gender discordance…. There have been no randomized controlled trials of any treatment….the proposed benefits of treatment to eliminate gender discordance…must be carefully weighed against… possible deleterious effects' (AACAP Practice Parameter on Gay, Lesbian, or Bisexual Sexual Orientation, Gender Nonconformity, and Gender Discordance in Children and Adolescents, 2012, pp. 968–969). Given the cautious conclusions that these types of reviews have reached, it is of interest how, in recent years, so many clinicians have embraced the treatment approach that recommends an early gender social transition. Chen, Edwards-Leeper, Stancin, and Tishelman (2018) observed that 'Over the last decade, we have seen a sea change in approach to pediatric transgender care, with the gender affirmative model now widely adopted as preferred practice' (p. 74).

In my view, there are reasons to be skeptical about the merit in recommending an early gender social transition as a first-line treatment. One should recognize that if one peruses carefully the follow-up studies of young children with gender dysphoria (or traits of gender dysphoria), the majority of such children do not have gender

Published by John Wiley & Sons Ltd, 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main St, Malden, MA 02148, USA

dysphoria when followed up in adolescence or adulthood (Zucker, 2018). In these studies, one can say with reasonable confidence that when these children had treatment (and not all did), the one type of treatment they did not receive was in the form of a prepubertal gender social transition. As I argued elsewhere (Zucker, 2018), if one conceptualizes gender social transition as a type of psychosocial treatment, it should come as no surprise that the rate of gender dysphoria persistence will be much higher as these children are followed into their adolescence and young adulthood (see Rae et al., 2019). If this is, in fact, the case, one might ask why would one recommend a first-line treatment that is, in effect, iatrogenic.

Even if there was a team of researchers motivated to design an RCT, the implementation of such a study would be formidable. For example, some parents would decline to place their child into a psychosocial treatment arm that would attempt to reduce the child's gender dysphoria so as to be more congruent with the gender assigned at birth; other parents would decline to place their child into a psychosocial treatment arm that would attempt to reduce the child's gender dysphoria by 'affirming' their felt gender vis-a-vis a social transition. Perhaps parents who prefer one of these two approaches would agree to 'wait-and-see' at least for a while, before deciding on a more intensive therapeutic approach. This variation in parental preferences reflects, as noted earlier, differences in underlying theoretical and philosophical perspectives which need to be respected. As the field moves forward and more follow-up data become available, we will learn more about the developmental course of gender dysphoria in particular and well-being and mental health in general.

Acknowledgements

The author has declared that he has no competing or potential conflict of interest.

Ethical information

No ethical approval was required for this article.

Correspondence

Kenneth J. Zucker, Department of Psychiatry, University of Toronto, Toronto, ON M5T 1R8, Canada; Email: ken.zucker@utoronto.ca

References

American Academy of Child and Adolescent Psychiatry. (2012). Practice parameter on gay, lesbian, or bisexual sexual orientation, gender nonconformity, and gender discordance in children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, *51*, 957–974.

American Psychological Association. (2015). Guidelines for psychological practice with transgender and gender nonconforming people. *American Psychologist*, *70*, 832–864.

Byne, W., Bradley, S.J., Coleman, E., Eyler, A.E., Green, R., Menvielle, E.J., ... & Tompkins, D.A. (2012). Report of the American Psychiatric Association Task Force on Treatment of Gender Identity Disorder. *Archives of Sexual Behavior*, *41*, 759–796.

Chen, D., Edwards-Leeper, L., Stancin, T., & Tishelman, A. (2018). Advancing the practice of pediatric psychology with transgender youth: State of the science, ongoing controversies, and future directions. *Clinical Practice in Pediatric Psychology*, *6*, 73–83.

Dreger, A. (2009). Gender identity disorder in childhood: Inconclusive advice to parents. *Hastings Center Report*, *39*, 26–29.

Drescher, J., & Byne, W. (2012). Introduction to the Special Issue on "The Treatment of Gender Dysphoria/Gender Variant Children and Adolescents". *Journal of Homosexuality*, *59*, 295–300.

Herthel, J., & Jennings, J. (2014). *I am Jazz*. New York: Penguin Group.

Rae, J.R., Gulgoz, S., Durwood, L., DeMeules, M., Lowe, R., Lindquist, G., & Olson, K.R. (2019). Predicting early-childhood gender transitions. *Psychological Science*, https://doi.org/10.1177/0956797619830649.

The Oxford Dictionary of Phrase and Fable. (2006). Oxford: Oxford University Press.

Zucker, K.J. (2018). The myth of persistence: Response to "A Critical Commentary on Follow-Up Studies and Desistance Theories about Transgender and Gender Non-Conforming Children" by Temple Newhook et al. (2018). *International Journal of Transgenderism*, *19*, 231–245.

Accepted for publication: 18 March 2019

Gender Identity 5 Years After Social Transition

Kristina R. Olson, PhD,[a] Lily Durwood, PhD,[b] Rachel Horton, BS,[a] Natalie M. Gallagher, PhD,[a] and Aaron Devor, PhD[c]

abstract

BACKGROUND AND OBJECTIVES: Concerns about early childhood social transitions among transgender youth include that these youth may later change their gender identification (ie, retransition), a process that could be distressing. The current study aimed to provide the first estimate of retransitioning and to report the current gender identities of youth an average of 5 years after their initial social transitions.

METHODS: The current study examined the rate of retransition and current gender identities of 317 initially transgender youth (208 transgender girls, 109 transgender boys; $M = 8.1$ years at start of study) participating in a longitudinal study, the Trans Youth Project. Data were reported by youth and their parents through in-person or online visits or via e-mail or phone correspondence.

RESULTS: We found that an average of 5 years after their initial social transition, 7.3% of youth had retransitioned at least once. At the end of this period, most youth identified as binary transgender youth (94%), including 1.3% who retransitioned to another identity before returning to their binary transgender identity. A total of 2.5% of youth identified as cisgender and 3.5% as nonbinary. Later cisgender identities were more common among youth whose initial social transition occurred before age 6 years; their retransitions often occurred before age 10 years.

CONCLUSIONS: These results suggest that retransitions are infrequent. More commonly, transgender youth who socially transitioned at early ages continued to identify that way. Nonetheless, understanding retransitions is crucial for clinicians and families to help make retransitions as smooth as possible for youth.

Increasing numbers of children are socially transitioning to live in line with their gender identity, rather than the gender assumed by their sex at birth, a process that typically involves changing a child's pronouns, first name, hairstyle, and clothing. Some concerns about childhood social transitions have been raised,[1] including that these children may not continue to identify as transgender, rather they might "retransition" (also called a "detransition" or "desistence"), which some suggest could be distressing for youth.[1–3] Research has suggested that ages 10 to 13 years may be particularly key times for retransition and that identity may be more stable after this period for youth who show early gender nonconformity.[3]

Other clinicians argue that early social transitions can be beneficial for some gender-diverse youth.[4–6] Some clinicians and scholars who support early childhood social transitions encourage families to remain open to later retransitions,[7,8] which are seen by some as part of a youth's exploration of their gender.[9]

Unfortunately, very few data about retransitions exist in the scientific literature. We have been able to find limited data on the number of youth

[a]*Department of Psychology, Princeton University, Princeton, New Jersey;* [b]*Department of Psychology, University of Washington, Seattle, Washington; and* [c]*Department of Sociology, University of Victoria, Victoria, BC, Canada*

Dr Olson conceptualized the current study, supervised data collection, carried out the initial analyses, and drafted the initial manuscript. Dr Durwood and Dr Devor conceptualized the current study and provided extensive revisions on the manuscript. Ms Horton acquired and compiled the data and tables and provided feedback on the manuscript. Dr Gallagher acquired, compiled, and analyzed the data and provided feedback on the manuscript. All authors approved the manuscript as submitted and agree to be accountable for all aspects of the work.

DOI: https://doi.org/10.1542/peds.2021-056082

Accepted for publication Apr 22, 2022

Address correspondence to: Kristina Olson, PhD, Department of Psychology, Princeton University, Peretsman Scully Hall, Princeton, NJ 08540. E-mail: krolson@princeton.edu

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

FUNDING: The collection of data in this report and some of the authors' time (Dr Olson, Ms Horton, and Dr Gallagher) over the course of this project have been supported by National Institutes of Health grant HD092347; National Science Foundation grants BCS-1523632, SMA-1837857, and SMA-2041463; an Arcus Foundation grant (all to Dr Olson); and a National Science Foundation GRFP grant to Dr Durwood. In addition, Dr Olson receives financial support from the MacArthur Foundation. The funders played no role in determining the research questions asked nor have they seen the results before review.

CONFLICT OF INTEREST DISCLOSURES: The authors have indicated that they have no potential conflicts of interest to disclose.

To cite: Olson KR, Durwood L, Horton R, et al. Gender Identity 5 Years After Social Transition. *Pediatrics*. 2022;150(2):e2021056082

who socially transition in childhood and then go on to retransition afterward. One paper included 4 youth who socially transitioned; none of them had retransitioned 7 years later.[10] We know of 3 mentions of early-transitioning youth who retransition.[8,9] However, these papers include no mention of how many other youth the same clinical team saw who did not retransition, making it impossible to guess a retransition rate.

In the present paper, we aimed to compute an estimate of retransition among a cohort of more than 300 early-transitioning children. Here, we report the retransition rate an average of 5 years after initial (binary) social transition, as well as how many of these participants are living as binary transgender youth, nonbinary youth, and cisgender youth at the same timepoint.

METHODS

A total of 317 binary socially transitioned transgender children (M_{age} = 8.07; SD = 2.36; 208 initially transgender girls, 109 initially transgender boys; see Table 1 for additional demographics) joined this longitudinal study (The Trans Youth Project) between July 2013 and December 2017. For inclusion in The Trans Youth Project, children had to be between 3 and 12 years of age and had to have made a "complete" binary social transition,[10] including changing their pronouns to the binary gender pronouns that differed from those used at their births.

As part of the larger longitudinal study, parents and youth were regularly asked about whether they had begun using puberty blockers and/or gender-affirming hormones. At most visits, they were not asked about whether puberty had begun, though our available data suggests that because these youth had socially transitioned at such early ages, most participants were followed by an endocrinologist well before puberty began. The endocrinologists helped families identify the onset of Tanner 2 (the first stage of puberty) and prescribed puberty blockers within a few months of this time; therefore, the onset of puberty blockers is used as our proxy for the onset of puberty in youth who received blockers. Of the youth in this sample, 37 (11.7%) had begun puberty blockers before beginning this study.

This study did not assess whether participants met criteria for the Diagnostic and Statistical Manual of Mental Disorders, Fifth edition, diagnosis of gender dysphoria in children. Many parents in this study did not believe that such diagnoses were either ethical or useful, even if they had been diagnosed, and some children did not experience the required distress criterion after transitioning. Based on data collected at their initial visit, these participants showed signs of gender identification and gender-typed preferences commonly associated with their gender, not their sex assigned at birth.[11] Further, parent report using the Gender Identity Questionnaire for Children[12] indicated that youth showed significant "cross-sex" identification and preferences (when scored based on sex at birth).[12]

Final identity classification for these analyses was based on our most recent interaction with the child and/or their parent before January 1, 2021. Because some families have not participated recently, we also separately report (Table 2) the results of the n = 291 youth with whom the research team had an interaction within the 2 years before that deadline. This additional analysis allows us to assess whether those who retransitioned were more likely to have missed their more recent appointments with our team. Importantly, only 1 of the 26 families with whom we did not meet in the past 2 years has formally dropped out of the study; the others often did not complete participation during these 2 years because of personal circumstances at the time we attempted re-recruitment. We anticipate that many in this group will participate again in the future.

TABLE 1 Participant Demographics (N = 317)

Demographics	%
Race	
White, non-Hispanic	69
White, Hispanic	9
Black	2
Asian	3
Native American	<1
Multiracial	17
Annual household income, $	
<25 000	3
25 001–50 000	10
50 001–75 000	21
75 001–125 000	31
>125 000	35
Location	
Northeast	15
Midwest/Upper Plains	21
Southeast	15
Mountain West	13
Pacific Northwest	20
Pacific South	16

Based on pronouns at follow-up, participants were classified as binary transgender (pronouns associated with the other binary assigned sex), nonbinary (they/them pronouns or, n = 3, a mix of they/them and binary pronouns), or cisgender (pronouns associated with their assigned sex). We confirmed this classification by reviewing other information available to the research team (eg, child's self-categorization in an interview or survey, e-mail communications with the parents). Only 1 classification was debatable; this participant was classified by pronouns (and in this paper) as nonbinary but could have been

TABLE 2 Participant Information and Current Identity at Last Visit Before January 1, 2021, Overall, for Those With Recent Visits Only, and by Initial Social Transition and Gender

	Total Sample	Recent Sample (With Visits in 2019 or 2020)	Sample Who Initially Socially Transitioned Before Age 6	Sample Who Initially Socially Transitioned at Age 6 or Later	Transgender Girls (At Recruitment)	Transgender Boys (At Recruitment)
Sample size	317	291	124	193	208	109
Assigned male at birth, %	65.6	65.3	73.4	60.6	100	0
Mean age at first transition, y	6.5	6.4	4.3	7.9	6.2	7.1
Mean age at start of study, y	8.1	8.0	5.9	9.5	7.7	8.7
Average time since start of study, y	3.8	4.1	3.8	3.8	3.9	3.7
Average time since first transition, y	5.4	5.7	5.4	5.4	5.5	5.3
Current identity, *n* (%)						
Binary transgender	298 (94.0)	276 (94.8)	112 (90.3)	186 (96.4)	194 (93.3)	104 (95.4)
Cisgender	8 (2.5)	6 (2.1)	7 (5.6)	1 (0.5)	7 (3.40)	1 (0.9)
Nonbinary	11 (3.5)	9 (3.1)	5 (4.0)	6 (3.1)	7 (3.40)	4 (3.7)

classified as binary transgender (and not retransitioned).

This study has been approved by the University of Washington and Princeton University institutional review boards.

RESULTS

The overall rate of retransition was 7.3%. An average of 5.37 years (SD = 1.74 years) after their initial binary social transition, most participants were living as binary transgender youth (94.0%; Table 2). Included in this group were 4 individuals (1.3% of the total sample) who retransitioned twice (to nonbinary then back to binary transgender). Some youth (3.5%) were currently living as nonbinary, including one who had retransitioned first to cisgender then to nonbinary. Finally, 2.5% were using pronouns associated with their sex at birth and could be categorized as cisgender at the time of data collection, including one who first retransitioned to live as nonbinary. Similar percentages were observed when examining the 291 youth who were in touch with the research team in the past 2 years (Table 2), when examining only those 280 youth who had not begun puberty blockers at the start of the study (Table 3), or if we examine only the 200 youth who had gone at least 5 years since their initial transition (Table 3).

We observed 1 potential (post hoc) age effect. Youth who initially socially transitioned before age 6 (n = 124), were more likely to be living as cisgender (n = 7; 5.6%) than youth who transitioned at age 6 or later (n = 1 of 193; 0.5%), Fisher exact test (comparing binary, cisgender, nonbinary; before vs. age 6 years or later), P = .02, although low rates of retransition were seen in both groups. In Table 2, we also report the results separately for children assigned male versus female at birth; this distinction was not significantly associated with later identity, P = .47, Fisher exact test. Finally, for exploratory purposes, in Table 3, we report outcomes separately for several subsets of our participants, including youth who had started puberty blockers, youth who had used puberty blockers and gender-affirming hormones, and youth who are at least 14 years old (the age at which past work[3] has suggested retransitions will be less likely).

DISCUSSION

Five years after an initial binary social transition, 7% of youth had retransitioned at least once. Most youth (94%) were living as binary transgender youth at the time of data analysis, including 1.3% who retransitioned initially to cisgender or nonbinary and then retransitioned back to binary trans identities. A small number of youth were living as cisgender youth (2.5%) or nonbinary youth (3.5%). We observed comparable rates when examining all participants who began the study (n = 317), those who had been in touch with the research team in the last two years (n = 291), those who had gone at least 5 years since initial social transition (n = 200), and

TABLE 3 Participant Information and Current Identity at Last Visit Before January 1, 2021, as a Function of Stages of Medical Transition and/or Age

	Total Sample	Sample of Youth Who Had Not Begun Blockers at Start of the Study	Sample of Youth Who Have Begun Blockers (and Not Gender-Affirming Hormones) at the End of the Study	Sample of Youth Who Have Begun Gender-Affirming Hormones at the End of the study	Sample of Youth 5+ y of Age Since Initial Binary Social Transition	Sample of Youth Who Are Currently 14+ y of Age
Sample size	317	280	92	98	200	70
Assigned male at birth, %	65.6	69.6	57.6	58.2	69.0	52.9
Mean age at first transition, y	6.5	6.1	6.6	8.4	6.2	8.9
Mean age at start of study, y	8.1	7.6	8.3	10.2	8.0	10.8
Average time since start of study, y	3.8	3.9	4	4.3	4.5	4.4
Average time since first transition	5.4	5.5	5.8	6.1	6.4	6.3
Current identity						
Binary transgender	$n = 298$; 94.0%	$n = 263$; 93.9%	$n = 88$; 95.7%	$n = 97$; 99.0%	$n = 190$; 95.0%	$n = 69$; 98.6%
Cisgender	$n = 8$; 2.5%	$n = 8$; 2.9%	$n = 1$; 1.1%	$n = 0$	$n = 4$; 2.0%	$n = 1$; 1.4%
Nonbinary	$n = 11$; 3.5%	$n = 9$; 3.2%	$n = 3$; 3.3%	$n = 1$, 1.0%	$n = 6$; 3.0%	$n = 0$

those who started the study before beginning puberty blockers ($n = 280$). We found no differences as a function of participant sex at birth. We observed slightly higher rates of retransition, and particularly later cisgender identity, among youth who initially socially transitioned before age 6 years. However, even in these youth, retransition rates were very low.

Among those who had begun puberty blockers and/or gender-affirming hormones, only 1 had retransitioned to live as cisgender (and this youth had begun blockers, but not gender-affirming hormones). One likely reason so few retransitions to cisgender occurred among those accessing medical transition is that most retransitioning in this cohort happened at early ages. All but 1 of the 8 cisgender youth had retransitioned by age 9 years (the last retransition was at age 11 years). Some of these youth are still not eligible for blockers because they are still prepubertal; we anticipate that those who identify as cisgender are unlikely to seek blockers or hormones, but that the participants who have not begun puberty and who identify as binary transgender or nonbinary likely will.

Past work has suggested that the ages 10 to 13 years are an especially critical time for retransition.[3] In our sample, many of the youth who retransitioned did so before that time frame, particularly the cisgender youth. In the nonbinary group, however, 6 of 11 retransitioned between ages 10 and 13 years, with the remainder retransitioning before age 10. Importantly, our sample differed from the past work on which this age range was determined in several key ways, including that our participants socially transitioned at earlier ages (perhaps pushing retransitions earlier, too), had undergone complete social transitions including pronouns and names (not just hairstyle and clothing changes as in most cases in previous studies[3]), and are living at a different historic time in a different country. Any, or all, of these may turn out to be key differences related to age of retransition.

Our observed low retransition rate is consistent with a study in which 4 youth who had completely socially transitioned had not retransitioned 7 years later.[10] That finding is in the same ballpark as our study's estimate of ~2.5% if we examine the percentage living as cisgender at the end of the study (ie, those "desisting" from gender-diverse outcomes). Together, these papers suggest this outcome is relatively rare in this group.

Our observation that few youth who have begun medical intervention have retransitioned to live as cisgender is consistent with findings in the literature. Several studies reporting on outcomes among transgender youth receiving blockers and gender-affirming hormones have reported relatively low rates of regret or stopping treatment,[13] which are potential indicators of retransition, though stopping treatment can occur for other reasons as well (eg, side

effects), as can regret (eg, experiences of transphobia).

Our key finding, that there was a relatively low rate of retransition about 5 years after initial social transition, may, on the surface, appear contradictory with past clinic-based research on what is sometimes called persistence and desistence[3] of childhood gender dysphoria. Several large studies attempted to recontact adolescents and adults who had previously been evaluated for gender dysphoria in childhood.[14–17] Many of those were formally diagnosed with what was, at the time, called gender identity disorder. Those studies reported that a minority of youth later identified in a way that might indicate a transgender identity by today's definition.

Interpretation of those results, and especially comparison with the present work, is difficult for several reasons. First, in past studies, when asked "are you a boy or a girl?" about 90% of the children supplied answers that aligned with their sex at birth,[18] leading some to question whether the majority of those children were the equivalent of transgender children today or not.[19–21] Second, participants in those studies were children between the 1960s and the 1990s, and many features of society have changed since then, including greater rates of acceptance and acknowledgment of transgender identities. Third, the parents of the youth in the current study support their children's identities, as indicated by their approval of their social transitions, whereas many of the parents of youth in past studies explicitly discouraged gender nonconformity or "cross-gender" identification.[15,22] In addition, it would have been exceedingly rare for youth in those studies to socially transition, especially completely.[1,10] Finally, there were substantial drop-out rates in all of the previous studies,[14,15,17] making the true estimates of persistence or desistence difficult to obtain.[19,21] Because there are so many possible contributors to differences in rates of persistence (in past work) and retransition in the current work, we urge caution about overinterpreting differences, or overconfidence about which contributing factors explain the differences.

There are also some reasons why we might have had such a low retransition rate. First, on average, participants had socially transitioned 1.6 years before joining our study. It is possible that some youth initially try socially transitioning and then change their minds quickly. Such youth would be unlikely to be enrolled in this study because their eligibility period would have been quite short and therefore the odds of finding the study and completing it would have been low. This means the children in our study may have been especially unlikely, compared with all children who transition, to retransition because they had already lived and presumably been fairly content with that initial transition for more than a year. Second, it is possible that families who failed to participate in the past 2 years of our study ($n = 26$) were disproportionately those whose children retransitioned and who were therefore hesitant to participate again. If true, their exclusion could have reduced our retransition rate. We are skeptical of this possibility for a few reasons. First, 4 of these participants did retransition and had told us about that outcome, so it does not appear that hesitancy in telling us was widespread in this group. Second, many of these families continue to be in touch with our research team and only missed participation because of ongoing personal issues (eg, COVID-19, emergency family circumstances). We anticipate that most of these families will be able to participate as we continue to follow these youth. Finally, from the beginning of the study, the research team has been clear in discussing with the families that we are open to any outcome in their youth.

As with past work, the present work has several key limitations. First, this is a volunteer community sample, meaning there could be biases in the kinds of families who sign up to participate. We know, for example, that unlike many samples of transgender youth, this sample of youth have normative levels of depression and only slight elevations in anxiety.[23] The parents of the participants in this study are disproportionately higher income and went to college at higher rates than the general population. We do not know whether these potential biases in the sample reflect biases in the cohort of children who socially transitioned in the mid-2010s in the United States and Canada. Therefore, whether the results generalize to youth without these characteristics is unknown.

Another potential limitation is that we used pronouns as the criterion for retransitions. Not everyone who, for example, uses they/them pronouns identifies as nonbinary and someone might identify as transgender even if they are currently using pronouns associated with their sex at birth. However, examination of other data provided by families suggests that our pronoun-based criteria were largely consistent with classification that would have arisen from other types of information provided to the research team (eg, labels used in an interview). Only 1 of the youth categorized as "retransitioned" might, by some other criteria, not meet that definition. However, because pronouns were the initial

inclusion criterion (that is, to be in the study children had to be using pronouns not associated with their sex at birth), they were the most consistent route of classification.

A related potential concern with these analyses is that we classified a change from using, for example, binary transgender to nonbinary as a retransition. Not everyone would categorize this change as a retransition. Many nonbinary people consider themselves to be transgender.[24] If we had used a stricter criterion of retransition, more similar to the common use of terms like detransition or desistence, referring only to youth who are living as cisgender, then our retransition rate would have been lower (2.5%).

One additional limitation in the present work is that the initial sample was disproportionately made up of trans girls. This is counter to recent reports that more peri- and postpubertal transgender youth seeking clinical services recently are transmasculine.[25–27] Historically, and consistent with our data, samples of parent-identified prepubertal gender nonconforming youth have included more assigned males at birth.[15,16,22] Importantly, we did not observe a significant gender effect in terms of rates of retransition, so we do not predict any change in pattern of results if we had a different ratio of participants by sex at birth.

We anticipate continuing to follow this cohort into adolescence and adulthood. This continued follow-up is necessary because it is possible that as more youth move into adolescence and adulthood, their identities could change. As we already saw, some youth will retransition more than once, so the present identities should not be interpreted as final.

As more youth are coming out and being supported in their transitions early in development, it is increasingly critical that clinicians understand the experiences of this cohort and not make assumptions about them as a function of older data from youth who lived under different circumstances. Though we can never predict the exact gender trajectory of any child, these data suggest that many youth who identify as transgender early, and are supported through a social transition, will continue to identify as transgender 5 years after initial social transition. These results also suggest that retransitions to one's gender assumed at birth (cisgender) might be likely to occur before age 10 years among those who socially transition at the earliest ages (before age 6 years), though retransitions are still unlikely in this group. These data suggest that parents and clinicians should be informed that not all youth will continue the same trajectory over time. Further understanding of how to support youth's initial and later transitions is needed.

ACKNOWLEDGMENTS

The authors thank the families for their participation in this research.

REFERENCES

1. Steensma TD, Cohen-Kettenis PT. Gender transitioning before puberty? *Arch Sex Behav*. 2011;40(4):649–650
2. de Vries ALC, Cohen-Kettenis PT. Clinical management of gender dysphoria in children and adolescents: the Dutch approach. *J Homosex*. 2012;59(3):301–320
3. Steensma TD, Biemond R, de Boer F, Cohen-Kettenis PT. Desisting and persisting gender dysphoria after childhood: a qualitative follow-up study. *Clin Child Psychol Psychiatry*. 2011; 16(4):499–516
4. Ashley F. Thinking an ethics of gender exploration: against delaying transition for transgender and gender creative youth. *Clin Child Psychol Psychiatry*. 2019;24(2):223–236
5. Sherer I. Social transition: supporting our youngest transgender children. *Pediatrics*. 2016;137(3):e20154358
6. Temple Newhook J, Pyne J, Winters K, et al. A critical commentary on follow-up studies and "desistance" theories about transgender and gender-nonconforming children. *Int J Transgenderism*. 2018; 19(2):212–224
7. Leibowitz S. Social gender transition and the psychological interventions. In: Janssen A, Leibowitz S, eds. *Affirmative Mental Health Care for Transgender and Gender Diverse Youth*. New York: Springer International Publishing; 2018:31–47
8. Edwards-Leeper L, Spack NP. Psychological evaluation and medical treatment of transgender youth in an interdisciplinary "Gender Management Service" (GeMS) in a major pediatric center. *J Homosex*. 2012;59(3):321–336
9. Menvielle E. A comprehensive program for children with gender variant behaviors and gender identity disorders. *J Homosex*. 2012;59(3):357–368
10. Steensma TD, McGuire JK, Kreukels BP, Beekman AJ, Cohen-Kettenis PT. Factors associated with desistence and persistence of childhood gender dysphoria: a quantitative follow-up study. *J Am Acad Child Adolesc Psychiatry*. 2013;52(6):582–590
11. Gülgöz S, Glazier JJ, Enright EA, et al. Similarity in transgender and cisgender children's gender development. *Proc Natl Acad Sci USA*. 2019;116(49):24480–24485
12. Johnson LL, Bradley SJ, Birkenfeld-Adams AS, et al. A parent-report gender identity questionnaire for children. *Arch Sex Behav*. 2004;33(2):105–116
13. Kuper LE, Stewart S, Preston S, Lau M, Lopez X. Body dissatisfaction and mental health outcomes of youth on gender-affirming hormone therapy. *Pediatrics*. 2020;145(4):e20193006
14. Drummond KD, Bradley SJ, Peterson-Badali M, Zucker KJ. A follow-up study of girls with gender identity disorder. *Dev Psychol*. 2008;44(1):34–45
15. Green R. *The Sissy Boy Syndrome: The Development of Homosexuality*. New Haven, CT: Yale University Press; 1987

16. Singh D, Bradley SJ, Zucker KJ. A follow-up study of boys with gender identity disorder. *Front Psychiatry*. 2021;12:632784

17. Wallien MS, Cohen-Kettenis PT. Psychosexual outcome of gender-dysphoric children. *J Am Acad Child Adolesc Psychiatry*. 2008; 47(12):1413–1423

18. Zucker KJ, Bradley SJ, Sullivan CB, Kuksis M, Birkenfeld-Adams A, Mitchell JN. A gender identity interview for children. *J Pers Assess*. 1993;61(3):443–456

19. Ashley F. The clinical irrelevance of "desistance" research for transgender and gender creative youth [published online ahead of print 2021]. *Psychol Sex Orientat Gend Divers*. 10.1037/sgd0000504

20. Olson KR. Prepubescent transgender children: what we do and do not know. *J Am Acad Child Adolesc Psychiatry*. 2016;55(3):155–156

21. Temple Newhook J, Pyne J, Winters K, et al. A critical commentary on follow-up studies and "desistance" theories about transgender and gender-nonconforming children. *Int J Transgenderism*. 2018;19(2):212–224

22. Zucker K, Bradley S. Gender identity disorder and psychosexual problems. In: *Children and Adolescents*. New York: Guilford Press; 1995

23. Gibson DJ, Glazier JJ, Olson KR. Evaluation of anxiety and depression in a community sample of transgender youth. *JAMA Netw Open*. 2021;4(4):e214739

24. Darwin H. Challenging the cisgender/transgender binary: nonbinary people and the transgender label. *Gend Soc*. 2020;34(3):357–380

25. Aitken M, Steensma T, Blanchard R, et al. Evidence for an altered sex ratio in clinic-referred adolescents with gender dysphoria. *J Sex Med*. 2015;12(3):756–763

26. de Graaf NM, Carmichael P, Steensma TD, Zucker KJ. Evidence for a change in the sex ratio of children referred for gender dysphoria: data from the Gender Identity Development Service in London (2000-2017). *J Sex Med*. 2018;15(10):1381–1383

27. Meyenburg B, Renter-Schmidt K, Schmidt G. Transidentität in Jugend und Adoleszenz: Zur Veränderung der Sexratio und der Prävalenz in den letzten eineinhalb Jahrzehnten [Changes of sex ratio and prevalence in transgender teenagers over the past 15 years]. *Z Kinder Jugendpsychiatr Psychother*. 2021;49(2):93–100

App.0525

Body Dissatisfaction and Mental Health Outcomes of Youth on Gender-Affirming Hormone Therapy

Laura E. Kuper, PhD,[a,b] Sunita Stewart, PhD,[a,b] Stephanie Preston, MD,[b] May Lau, MD, MPH,[a,b] Ximena Lopez, MD[a,b]

abstract

OBJECTIVES: Our first aim was to examine baseline differences in body dissatisfaction, depression, and anxiety symptoms by gender, age, and Tanner (ie, pubertal) stage. Our second aim was to test for changes in youth symptoms over the first year of receiving gender-affirming hormone therapy. Our third aim was to examine potential differences in change over time by demographic and treatment characteristics. Youth experiences of suicidal ideation, suicide attempt, and nonsuicidal self-injury (NSSI) are also reported.

METHODS: Participants (n = 148; ages 9–18 years; mean age 14.9 years) were receiving gender-affirming hormone therapy at a multidisciplinary program in Dallas, Texas (n = 25 puberty suppression only; n = 123 feminizing or masculinizing hormone therapy). Participants completed surveys assessing body dissatisfaction (Body Image Scale), depression (Quick Inventory of Depressive Symptoms), and anxiety (Screen for Child Anxiety Related Emotional Disorders) at initial presentation to the clinic and at follow-up. Clinicians completed the Quick Inventory of Depressive Symptoms and collected information on youth experiences of suicidal ideation, suicide attempt, and NSSI.

RESULTS: Affirmed males reported greater depression and anxiety at baseline, but these differences were small ($P < .01$). Youth reported large improvements in body dissatisfaction ($P < .001$), small to moderate improvements in self-report of depressive symptoms ($P < .001$), and small improvements in total anxiety symptoms ($P < .01$). No demographic or treatment-related characteristics were associated with change over time. Lifetime and follow-up rates were 81% and 39% for suicidal ideation, 16% and 4% for suicide attempt, and 52% and 18% for NSSI, respectively.

CONCLUSIONS: Results provide further evidence of the critical role of gender-affirming hormone therapy in reducing body dissatisfaction. Modest initial improvements in mental health were also evident.

[a]*Children's Health Systems of Texas, Dallas, Texas; and* [b]*University of Texas Southwestern Medical Center, Dallas, Texas*

Dr Kuper oversaw data collection, conducted data analysis, and drafted the manuscript; Drs Stewart, Lau, and Lopez conceptualized and designed the study and provided feedback on manuscript drafts; Dr Preston assisted with drafting the manuscript; and all authors contributed to the development of study aims, approved the final manuscript as submitted, and agree to be accountable for all aspects of the work.

DOI: https://doi.org/10.1542/peds.2019-3006

Accepted for publication Dec 6, 2019

Address correspondence to Laura E. Kuper, PhD, Department of Endocrinology, Children's Health Systems of Texas, 1935 Medical District Dr, Mail Code F4.05, Dallas, TX 75235. E-mail: laura.kuper@childrens.com

WHAT'S KNOWN ON THIS SUBJECT: Guidelines exist for providing gender-affirming hormone therapy (ie, puberty suppression and masculinizing or feminizing hormone therapy) to transgender youth; however, little research has been conducted on the impact of treatment on body dissatisfaction and mental health and factors that may influence this impact.

WHAT THIS STUDY ADDS: One year of receiving gender-affirming hormone therapy resulted in large reductions in youth body dissatisfaction and modest improvements in mental health. No demographic or treatment-related factors were associated with change over time.

To cite: Kuper LE, Stewart S, Preston S, et al. Body Dissatisfaction and Mental Health Outcomes of Youth on Gender-Affirming Hormone Therapy. *Pediatrics.* 2020; 145(4):e20193006

Two influential longitudinal studies from the Netherlands have helped establish guidelines for providing gender-affirming hormone therapy (ie, puberty suppression and masculinizing or feminizing hormone therapy) to transgender youth with gender dysphoria.[1,2] De Vries et al[3] conducted a prospective study with 70 youth who received puberty suppression (ie, medication to stop the progression of puberty). After 2 years, internalizing, externalizing, and depressive symptoms improved along with global functioning, but there was no improvement in body dissatisfaction or anxiety symptoms. A subset of the same cohort (n = 55) was reassessed after masculinizing or feminizing hormone therapy and gender-affirming surgery (vaginoplasty or mastectomy and hysterectomy), at which point there was a sustained improvement in global functioning and most measures of mental health. Gender dysphoria and body dissatisfaction also improved, and self-reported quality of life was similar to the Dutch population.[4] However, patients were not evaluated after masculinizing or feminizing hormone therapy alone.

In the only other longitudinal study of youth, participants seen in a gender clinic in the United Kingdom (n = 35) demonstrated improvement in clinician assessment of psychosocial functioning after 12 months of receiving puberty suppression.[5] Only 1 cross-sectional study has included a subset of transgender youth (n = 82 of 202). In comparison with those who had not started treatment, individuals who received both puberty suppression and/or masculinizing or feminizing hormone therapy as well as surgery had more favorable body image but not those who received puberty suppression and/or masculinizing or feminizing hormone therapy only.[6] Within this study, youth and adults as well as those receiving puberty suppression and/or masculinizing or feminizing hormone therapy were combined.

The benefits of gender-affirming treatment are better described in adults. A recent review of 5 longitudinal and 2 cross-sectional studies found that receipt of masculinizing or feminizing hormone therapy alone was associated with improved depression in 5 of 7 studies, improved anxiety in 2 of 2 studies, and better quality of life in 3 of 3 studies.[7] Two studies also found lower rates of body uneasiness in adults who received masculinizing or feminizing hormone therapy alone (ie, dissatisfaction with body parts and negative body-related experiences, such as avoidance and self-monitoring).[8,9]

Understanding the impact of gender-affirming hormone therapy on the mental health of transgender youth is critical given the health disparities documented in this population. Within samples of transgender youth presenting for gender-affirming hormone therapy, estimates of clinically significant depressive symptoms or diagnoses have averaged in the range of 30% to 60%,[10–13] and estimates of clinically significant anxiety symptoms or diagnoses have averaged in the range of 20% to 30%.[11,14–16] Lifetime history of suicidal ideation (average range 30%–50%),[10,11,16] suicide attempt (average range 15%–30%),[10,11,13] and nonsuicidal self-injury (NSSI) (average range 20%–40%)[12,13,16] also appear common.

There is also some evidence that rates of mental health concerns may vary by gender, but no clear pattern has emerged.[11,14,15,17] Two studies have found higher levels of body dissatisfaction among affirmed females (ie, individuals assigned male at birth who identify as female) in comparison with affirmed males (ie, individuals assigned female at birth who identify as male).[6,18] Changes associated with puberty, as reflected in age and/or Tanner stage (ie, stage of puberty), may exacerbate body dissatisfaction and mental health concerns. Fewer studies have examined differences by age; however, one study found greater symptoms of depression but not anxiety among older adolescents,[16] and one study found higher levels of body dissatisfaction.[4] None have specifically examined the impact of Tanner stage.

Our first aim in this study was to explore how transgender youth baseline body dissatisfaction, depression, and anxiety symptoms vary on the basis of their gender, age at initial assessment, and Tanner stage at first medical visit. Consistent with our earlier article examining differences in mental health functioning using the Child Behavior Checklist and Youth Self-Report,[14] we hypothesized that affirmed males will report greater symptoms of depression and anxiety. We also hypothesized that older age and greater Tanner stage will be associated with higher ratings of body dissatisfaction and more symptoms of depression and anxiety.

Our second aim was to examine how transgender youth body dissatisfaction, depression, and anxiety symptoms change over the first year of receiving gender-affirming hormone therapy. We anticipated improvements in each of these domains but did not have any a priori hypotheses regarding which domains would demonstrate the greatest improvements.

Our third aim was to explore how any changes over time vary by affirmed gender, Tanner stage, age, type of treatment, months on masculinizing or feminizing hormone therapy, mental health treatment received, and whether chest (ie, "top") surgery was also obtained (among those assigned female at birth). We hypothesized that older age, greater Tanner stage,

receipt of puberty suppression only, fewer months on masculinizing or feminizing hormone therapy, and lack of chest surgery will be associated with fewer changes over time. Lastly, for descriptive purposes, we report information on lifetime and follow-up rates of suicidal ideation, suicide attempts, NSSI, and mental health treatment.

METHODS

Participants and Procedure

Participants are youth who received gender-affirming hormone therapy with a multidisciplinary program in Dallas, Texas. Before initiating care, participants and their families participated in an initial assessment with the program's psychologist, psychiatrist, and/or clinical therapist after parents completed a phone intake survey and provided a referral letter from a licensed therapist or counselor documenting the presence of gender dysphoria (this letter is no longer required). Approximately 34% of families did not follow-up after the phone intake. Initial assessments occurred between August 2014 and March 2018, with most occurring in 2017 (41%) or 2016 (37%). At home before this visit, participants completed self-report measures of depression, anxiety, and body dissatisfaction. During the visit, clinicians also completed a report of depressive symptoms and collected information regarding lifetime and recent suicidal ideation, suicide attempts, and NSSI as well as current participation in therapy and support groups and use of psychiatric medication(s).

After the assessment, participants were discussed by the multidisciplinary team of providers from psychology, social work, pediatric endocrinology, pediatric and adolescent gynecology, and adolescent medicine. The Endocrine Society Clinical Practice Guidelines[2] guided the initiation of hormone therapy. Chest surgery was not performed within the program, but participants were provided with referrals when requested.

Approximately 1 year after this initial assessment (range: 11–18 months), all patients were asked to participate in a yearly reassessment visit. Participants were readministered self-report measures, and clinicians again completed a report of depressive symptoms and documented information about suicidal ideation, suicide attempts, NSSI, and mental health treatment.

Survey and clinician data were entered into a research database for analysis along with demographic and treatment-related information (ie, Tanner stage at first medical visit, treatment start and end dates, and chest surgery date extracted from physicians' notes). All participants provided consent, or assent with parent consent, to allow this information to be used for research. The study was approved by the institutional review board at the University of Texas Southwestern Medical Center.

Measures

Participants were asked to self-report their gender identity (all ages) and sexual orientation (age 12 and older). These responses were recorded verbatim by the clinician and entered into the research database. Gender identities were coded into the following categories: (1) male, boy, or man; (2) male spectrum (eg, "trans masculine" or "masculine nonbinary"); (3) female, girl, or woman; (4) female spectrum (eg, "mostly female, slightly nonbinary"); and (5) nonbinary (eg, "agender" or "part girl, part boy").

To assess body dissatisfaction, participants aged 12 years and older rated their degree of dissatisfaction with 29 areas of the body using the Body Image Scale (BIS).[19] Participants of all ages completed the Screen for Child Anxiety Related Emotional Disorders (SCARED), which produces a total score as well as subscale scores for panic-related, social, separation-related, generalized, and school avoidance–related anxiety symptoms,[20] as well as the Quick Inventory of Depressive Symptoms (QIDS)[21] to measure symptoms of depression that reflect the *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* criteria for major depressive disorder.[22] The QIDS produces a total score that can also be grouped into clinical categories: not elevated (0–5), mild (6–10), moderate (11–15), and severe (16–27). Clinicians also completed the clinician version of the QIDS. When the percentage of missing values for each total score and subscale score was ≤15%, missing values were imputed by using the mean of nonmissing values.

Analyses

To examine baseline differences in depression (QIDS self and clinician), anxiety (SCARED), and body dissatisfaction (BIS), bivariate correlation coefficients were first examined by using Pearson's r for age, Spearman's ρ for Tanner stage, and point biserial for gender. Variables with significant correlations were then simultaneously entered into a linear regression for each outcome, and Cohen's f^2 was calculated as a measure of effect size (0.1 = small, 0.25 = moderate, and 0.4 = large).[23]

To examine change over time, QIDS (self and clinician), SCARED, and BIS scores were first tested for normality by using the Kolmogorov-Smirnov test. Changes in normally distributed variables were examined by using paired *t* tests, and the Wilcoxon rank test was used when the Kolmogorov-Smirnov value was significant. Cohen's *d* was used as a measure of effect size (0.2 = small, 0.5 = moderate, and 0.8 = large).[23] Changes

in clinical groupings on the QIDS were also examined by using the Wilcoxon rank test. For both baseline and longitudinal analyses, we planned to first examine the SCARED total score then test for differences in subscale scores only if this change was significant.

To test for associations between change scores and demographic and treatment characteristics, change scores were calculated by subtracting baseline scores from follow-up scores for variables that exhibited a significant change over time. Bivariate correlation coefficients were then examined by using Pearson's r for age and months on feminizing or masculinizing hormone therapy, Spearman's ρ for Tanner stage and therapy frequency, and point biserial for gender, treatment type, psychiatric medication use, support group participation, and chest surgery receipt (for those assigned female at birth). We planned to include any variables with significant correlations in a linear regression. $P < .01$ was significant for all statistical tests to help account for the overall number of tests. Confidence intervals (CIs) are reported at the 95% level.

RESULTS

Figure 1 presents a flow diagram of participants who were due for follow-up (≥18 months since initial assessment), participants with follow-up data, and the reasons why follow-up data were not available or excluded. The mean number of months between initial assessment and reassessments was 14.9 (SD 2.1). Table 1 presents demographic information on participants. At the initial assessment, patients ranged in age from 9 to 18 years (mean 15.4; SD 2.0). All but 1 participant met *Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition* criteria for gender dysphoria. This participant

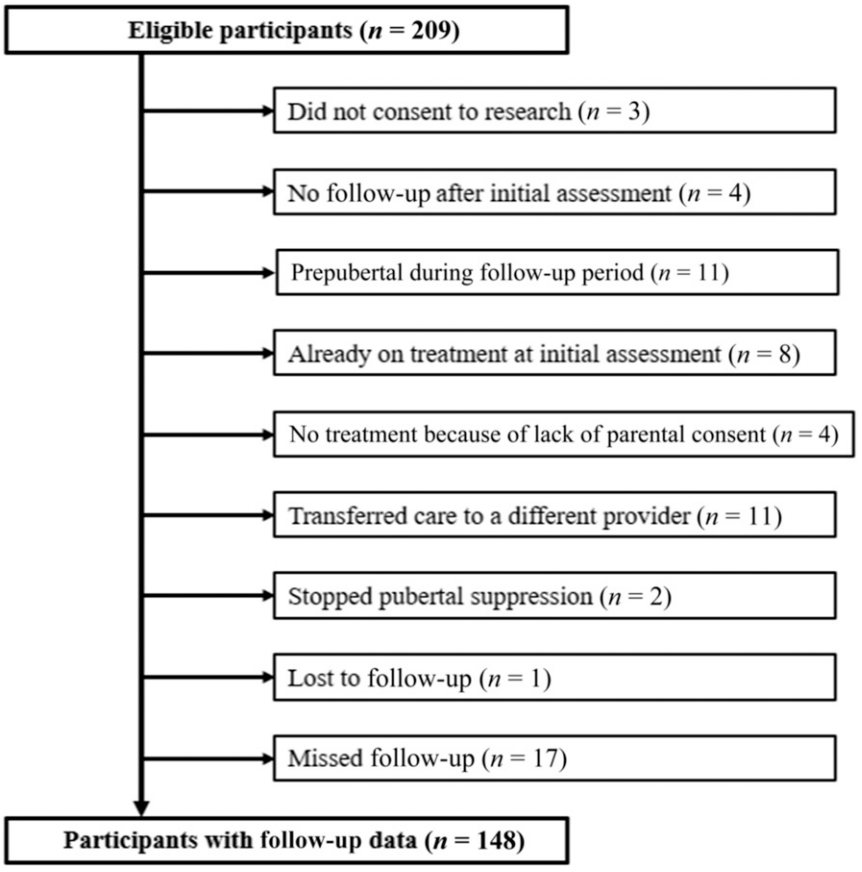


FIGURE 1
Flow diagram.

subsequently met criteria at a follow-up visit and was started on treatment. Participants who started puberty suppression only did so at a mean age of 13.7 years (range 9.8–14.9; SD 1.5), and participants started feminizing or masculinizing hormone therapy at a mean age of 16.2 years (range 13.2–18.6; SD 1.2). For participants who were on masculinizing or feminizing hormone therapy, the mean length of time receiving treatment before follow-up was 10.9 months (range 1–18; SD 3.3). During the follow-up period, 2 participants stopped puberty suppression without starting masculinizing or feminizing hormone therapy, and no participants stopped masculinizing or feminizing hormone therapy. Fifteen affirmed males obtained chest surgery at an average age of 17.1 years (range 15.2–18.7; SD 1.2) and at an average of 9.2 months from baseline (range 3.0–16.0; SD 3.3).

Table 2 presents means, SDs, and ranges for QIDS, SCARED, and BIS scores at initial assessment and follow-up for the full sample as well as by gender and treatment type. At baseline, affirmed males had greater clinician-reported depressive symptoms (CI −3.76 to −0.81), self-reported depressive symptoms (CI −4.46 to −0.79), total anxiety symptoms (CI −14.94 to −3.99), panic symptoms (CI −5.88 to −1.78), and school avoidance symptoms (CI −1.81, to −0.36) in comparison with affirmed females. However, Cohen's f^2 effect sizes were all in the small range (0.07, 0.06, 0.09, 0.10, and 0.07, respectively). No differences were found by age or Tanner stage.

Within the full sample, a significant decrease in body dissatisfaction (CI

TABLE 1 Participant Demographics

	n (%)
Gender identity	
Male, boy, or guy	81 (55)
Male spectrum	9 (6)
Female, girl, or woman	52 (35)
Female spectrum	2 (1)
Something else[a]	3 (2)
Assigned sex	
Male	55 (37)
Female	94 (63)
Sexual orientation[b]	
Pansexual	25 (20)
Straight	24 (19)
Bisexual	15 (12)
Gay	12 (10)
Unsure	12 (10)
No label	11 (9)
Asexual	10 (8)
Something else	10 (8)
Lesbian	6 (5)
Race	
White	137 (95)
African American	3 (2)
Multiracial	3 (2)
American Indian	1 (1)
Ethnicity	
Hispanic	24 (17)
Non-Hispanic	120 (83)
Tanner stage	
I	3 (2)
II	6 (4)
III	5 (4)
IV	32 (23)
V	94 (67)
Treatment type[c]	
Puberty suppression only	25 (17)
Masculinizing or femininizing therapy only	93 (63)
Both treatments	30 (20)

[a] Excluded from gender analyses.
[b] Age 12 and older.
[c] Masculinizing or feminizing therapy only and both treatments were collapsed for analysis by treatment type.

14.74 to 21.90), self-reported depressive symptoms (CI 1.24 to 2.97), and total anxiety symptoms (CI 1.05 to 6.70) was observed during the follow-up period. Decreases in generalized, separation, and school-related anxiety symptoms were significant at the $P < .05$ level but not the $P < .01$ level. No change in clinician report of depressive symptoms was found. Cohen's *d* effect sizes were large for change in BIS scores (1.04), small to moderate for change in QIDS self-report scores (0.44), and small for change in SCARED total scores (0.27). Table 3 reports the percentage of the sample that fell into each clinical category on the QIDS at initial assessment and follow-up. A significant change was also found in self-reported depressive symptom categories ($P < .001$) but not clinician-reported categories. No correlations were found between change scores and demographic and treatment-related characteristics. Although change scores were generally higher for participants who received chest surgery, no correlations were significant.

Table 4 presents descriptive data on mental health treatment, and Table 5 presents data on suicidal ideation, suicide attempt, and NSSI. During the follow-up period, the distribution of therapy frequency was as follows: none (16%), less than every 3 months (15%), every 2 to 3 months (12%), monthly (22%), every other week (21%), and weekly (14%). Of those who experienced suicidal ideation during the follow-up period, 94% had a lifetime history. These figures were 67% for suicide attempt and 87% for NSSI.

DISCUSSION

Youth reported large improvements in body dissatisfaction during the 1-year follow-up period. The amount of improvement was not related to treatment type. These findings are consistent with a handful of studies that have documented improvements in body dissatisfaction within samples of adults receiving feminizing or masculinizing hormone therapy[8,9] but contrast with the 2 existing studies of youth. Within the longitudinal cohort from Amsterdam, puberty suppression alone was not associated with improvements in body dissatisfaction,[3] and within a cross-sectional study with a mixed sample of youth and adults, puberty suppression and/or feminizing or masculinizing hormone therapy was not associated with more favorable body image.[6] In contrast to the Amsterdam sample, youth in the current study were younger when starting puberty suppression (age: mean 12.5 and range 9.8–14.9 versus mean 13.7 and range 11.1–17.0).

Age, puberty stage, length of time receiving feminizing or masculinizing hormone therapy, and receipt of chest surgery were also not associated with amount of improvement. However, the sample size of participants receiving puberty suppression only and chest surgery were small, and variations in months on feminizing or masculinizing hormone therapy may not have been meaningful enough in the relatively short follow-up period.

TABLE 2 Body Dissatisfaction, Depression, and Anxiety Symptoms at Baseline and Follow-up

	n	Range[a]	Baseline, Mean (SD)	Follow-up, Mean (SD)
Body dissatisfaction (BIS)		0–116		
Full sample[b]	96		69.9 (15.6)	51.7 (18.4)
Affirmed males	66		71.1 (13.4)	52.9 (16.8)
Affirmed females	30		67.5 (19.5)	49.0 (21.6)
Puberty suppression	10		64.1 (18.2)	53.8 (20.1)
Feminine or masculine hormone therapy	86		70.7 (15.2)	51.4 (18.3)
Depressive symptoms (QIDS), self report[c]		0–27		
Full sample[b]	118		9.4 (5.2)	7.3 (4.6)
Affirmed males	76		10.4 (5.0)	7.5 (4.5)
Affirmed females	40		7.5 (4.9)	6.6 (4.4)
Puberty suppression	13		8.2 (6.1)	7.0 (5.6)
Feminine or masculine hormone therapy	105		9.6 (5.0)	7.4 (4.5)
Depressive symptoms (QIDS), clinician report[c]		0–27		
Full sample	125		5.8 (4.2)	5.9 (3.9)
Affirmed males	78		6.7 (4.4)	6.2 (4.1)
Affirmed females	45		4.2 (3.2)	5.4 (3.4)
Puberty suppression	19		5.3 (4.9)	5.5 (4.8)
Feminine or masculine hormone therapy	106		5.9 (4.1)	6.0 (3.8)
Anxiety symptoms (SCARED), total score[c]		0–82		
Full sample[d]	102		32.4 (16.3)	28.6 (16.1)
Affirmed males	65		35.4 (16.5)	29.8 (15.5)
Affirmed females	33		26.4 (14.2)	24.3 (15.4)
Puberty suppression	22		31.8 (16.6)	29.3 (17.1)
Feminine or masculine hormone therapy	80		32.6 (16.3)	28.4 (15.9)
Panic symptoms (SCARED)[c]		0–26		
Full sample	104		8.2 (6.3)	7.1 (6.3)
Affirmed males	66		9.3 (6.5)	7.9 (6.5)
Affirmed females	34		5.7 (4.9)	5.1 (4.9)
Puberty suppression	22		8.7 (6.5)	7.2 (5.7)
Feminine or masculine hormone therapy	82		8.1 (6.3)	7.1 (6.5)
Generalized anxiety symptoms (SCARED)		0–18		
Full sample	104		9.7 (5.1)	8.7 (5.1)
Affirmed males	66		10.4 (5.0)	9.0 (5.1)
Affirmed females	34		8.6 (5.1)	8.0 (5.1)
Puberty suppression	22		8.5 (5.2)	8.2 (5.4)
Feminine or masculine hormone therapy	82		10.0 (5.1)	8.8 (5.0)
Social anxiety symptoms (SCARED)		0–14		
Full sample	104		8.0 (4.1)	7.6 (4.3)
Affirmed males	66		8.5 (4.0)	7.8 (4.1)
Affirmed females	34		7.1 (3.9)	6.8 (4.4)
Puberty suppression	22		6.3 (3.6)	7.3 (4.7)
Feminine or masculine hormone therapy	82		8.5 (4.1)	7.7 (4.2)
Separation anxiety symptoms (SCARED)[e]		0–16		
Full sample	103		4.0 (3.4)	3.3 (2.7)
Affirmed males	65		4.2 (3.4)	3.4 (2.6)
Affirmed females	34		3.4 (3.3)	2.7 (2.3)
Puberty suppression	22		5.8 (4.0)	4.2 (3.1)
Feminine or masculine hormone therapy	81		3.5 (3.0)	3.1 (2.5)
School avoidance symptoms (SCARED)[c]		0–8		
Full sample	102		2.6 (2.2)	2.0 (2.1)
Affirmed males	65		2.9 (2.3)	2.0 (2.3)
Affirmed females	33		1.8 (1.7)	1.9 (2.1)
Puberty suppression	22		2.6 (2.7)	2.4 (2.4)
Feminine or masculine hormone therapy	80		2.6 (2.1)	2.0 (2.0)

[a] Absolute range.
[b] Significant change from initial assessment to follow-up ($P < .001$).
[c] Significant difference in baseline scores by gender ($P < .01$).
[d] Significant change from initial assessment to follow-up ($P < .01$).
[e] Significant difference in baseline scores by age ($P < .01$).

TABLE 3 Depressive Symptoms (QIDS) Scoring Ranges

	Range	Self-Report[a]		Clinician Report	
		Baseline, *N* (%)	Follow-up, *N* (%)	Baseline, *N* (%)	Follow-up, *N* (%)
Not elevated	0–5	33 (25)	51 (40)	73 (53)	67 (49)
Mild	6–10	46 (35)	48 (37)	44 (32)	49 (36)
Moderate	11–15	29 (22)	22 (17)	15 (11)	16 (12)
Severe	16–27	24 (18)	8 (6)	5 (4)	4 (3)

[a] Significant change from initial assessment to follow-up ($P < .001$).

Most participants (90%) were also in advanced stages of puberty (Tanner stage IV or V) when presenting for care. Limitations associated with collecting data within a busy clinical setting with multiple providers also resulted in missing data. Nonetheless, results suggest that youth receiving gender-affirming hormone therapy experience meaningful short-term improvements in body dissatisfaction, and no participants discontinued feminizing or masculinizing hormone therapy. These results provide additional support for the incorporation of these treatments into the standards of care for transgender youth experiencing gender dysphoria.[1,2]

Youth also reported modest improvements in mental health functioning during the follow-up period. These results are consistent with the existing longitudinal studies of youth.[3–5] Several factors may help explain why improvements were not greater than what was observed. Although physical changes associated with feminizing or masculinizing hormone therapy often start within the first 3 months, changes continue over the course of several years. Furthermore, environmental stressors associated with one's transgender status may not improve after hormone therapy and could potentially worsen should they increase the youth's visibility as a transgender person. Research has consistently documented higher rates of bullying among transgender youth in comparison with nontransgender youth.[24,25] Within the current study, rates of school avoidance–related anxiety did not improve over the follow-up period.

The larger political context is also important to consider. Within Texas, where the current study was conducted, a well-publicized "bathroom bill" was introduced during the study period that prohibited transgender people from using a restroom that was different from the sex on their birth certificate, although the bill ultimately failed to pass.[26] As a whole, the mental health functioning of youth from the present clinic as well as youth from a handful of other US- and European-based clinics appears poorer than the mental health functioning of youth from the Amsterdam clinic.[11,14,17] Previous studies have attributed this difference to Amsterdam's social and political climate, which is known to be more supportive of the lesbian, gay, bisexual, and transgender population.[17]

TABLE 4 Mental Health Treatment

	At Initial Assessment, *n* (%)	Follow-up Period, *n* (%)
Psychiatric medication	67 (47)	80 (61)
Therapist or counselor	144 (97)	114 (84)
Support group[a]	60 (43)	45 (35)

[a] Participation by parents and/or youth (eg, transgender family support organization; lesbian, gay, bisexual, and transgender youth center; or school-based Gay-Straight Alliance).

Consistent with our study examining baseline differences in mental health functioning as measured by the Child Behavior Checklist and Youth Self-Report,[14] affirmed males reported greater symptoms of depression and several forms of anxiety in comparison with affirmed females. However, the effect size of these differences was smaller within the current study in comparison with the former. Differences in measurement approach may help explain the mixed findings regarding gender differences in mental health functioning across youth clinics.[11,15,17] Although some research suggests that nonclinic samples of affirmed male youth report more experiences of bullying,[24] affirmed females are thought to experience greater stigma regarding expression of femininity. Consistent with the current sample, the sex ratio of youth presenting to clinics also appears to be shifting from more affirmed females to more affirmed males presenting for care.[27] Although causes of this shift are largely unknown, they may be associated with other shifts in clinical presentations (eg, mental health and psychosocial functioning).

CONCLUSIONS

The current study is the largest longitudinal study of youth receiving gender-affirming hormone therapy to date and documents important improvements in body dissatisfaction over the first year of treatment. Continued longitudinal study of this

TABLE 5 Suicidal Ideation, Suicide Attempt, and NSSI

	Lifetime, *n* (%)	1–3 mo Before Initial Assessment,[a] *n* (%)	Follow-up Period, *n* (%)
Passive ideation	105 (81)	33 (25)	51 (38)
Suicide attempt	20 (15)	3 (2)	6 (5)
NSSI	68 (52)	13 (10)	23 (17)

[a] One month for passive ideation and 3 months for NSSI and suicide attempt(s).

population will increase the field's understanding of the benefits of gender-affirming hormone therapy and assist providers in better anticipating needs. Follow-up periods of several years or more will help document the full impact of the physical changes with feminizing or masculinizing hormone therapy, and larger sample sizes will improve the ability to examine the specific impacts of treatment type and chest surgery. Greater consideration of intersectionality and sociocultural context will further strengthen these efforts.

ACKNOWLEDGMENT

Rong Huang, MS, provided consultation on the data analysis.

ABBREVIATIONS

BIS: Body Image Scale
CI: confidence interval
NSSI: nonsuicidal self-injury
QIDS: Quick Inventory of Depressive Symptoms
SCARED: Screen for Child Anxiety Related Emotional Disorders

PEDIATRICS (ISSN Numbers: Print, 0031-4005; Online, 1098-4275).

FINANCIAL DISCLOSURE: The authors have indicated they have no financial relationships relevant to this article to disclose.

FUNDING: Research reported in this publication was supported by Children's Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of Children's Health. The Research Electronic Data Capture database was funded by the Clinical and Translational Science Awards program National Insitutes of Health grant UL1-RR024982 awarded to the UT Southwestern Center For Translational Medicine.

POTENTIAL CONFLICT OF INTEREST: Dr Lopez has participated as a member of an advisory board for Endo International; the other authors have indicated they have no potential conflicts of interest to disclose.

REFERENCES

1. Coleman E, Bockting W, Botzer M, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgenderism*. 2012;13(4):165–232
2. Hembree WC, Cohen-Kettenis PT, Gooren L, et al. Endocrine treatment of gender-dysphoric/gender-incongruent persons: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2017;102(11):3869–3903
3. de Vries AL, Steensma TD, Doreleijers TA, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. *J Sex Med*. 2011;8(8):2276–2283
4. de Vries AL, McGuire JK, Steensma TD, Wagenaar EC, Doreleijers TA, Cohen-Kettenis PT. Young adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*. 2014;134(4):696–704
5. Costa R, Dunsford M, Skagerberg E, Holt V, Carmichael P, Colizzi M. Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. *J Sex Med*. 2015;12(11):2206–2214
6. Becker I, Auer M, Barkmann C, et al. A cross-sectional multicenter study of multidimensional body image in adolescents and adults with gender dysphoria before and after transition-related medical interventions. *Arch Sex Behav*. 2018;47(8):2335–2347
7. Rowniak S, Bolt L, Sharifi C. Effect of cross-sex hormones on the quality of life, depression and anxiety of transgender individuals: a quantitative systematic review. *JBI Database Syst Rev Implement Rep*. 2019;17(9):1826–1854
8. Turan Ş, Aksoy Poyraz C, Usta Sağlam NG, et al. Alterations in body uneasiness, eating attitudes, and psychopathology before and after cross-sex hormonal treatment in patients with female-to-male gender dysphoria. *Arch Sex Behav*. 2018;47(8):2349–2361
9. Fisher AD, Castellini G, Ristori J, et al. Cross-sex hormone treatment and psychobiological changes in transsexual persons: two-year follow-up data. *J Clin Endocrinol Metab*. 2016;101(11):4260–4269
10. Olson J, Schrager SM, Belzer M, Simons LK, Clark LF. Baseline physiologic and psychosocial characteristics of transgender youth seeking care for gender dysphoria. *J Adolesc Health*. 2015;57(4):374–380
11. Reisner SL, Vetters R, Leclerc M, et al. Mental health of transgender youth in care at an adolescent urban community health center: a matched retrospective cohort study. *J Adolesc Health*. 2015;56(3):274–279
12. Spack NP, Edwards-Leeper L, Feldman HA, et al. Children and adolescents with gender identity disorder referred to

a pediatric medical center. *Pediatrics*. 2012;129(3):418–425

13. Peterson CM, Matthews A, Copps-Smith E, Conard LA. Suicidality, self-harm, and body dissatisfaction in transgender adolescents and emerging adults with gender dysphoria. *Suicide Life Threat Behav*. 2017;47(4):475–482
14. Kuper LE, Mathews S, Lau M. Baseline mental health and psychosocial functioning of transgender adolescents seeking gender-affirming hormone therapy. *J Dev Behav Pediatr*. 2019; 40(8):589–596
15. de Vries AL, Doreleijers TA, Steensma TD, Cohen-Kettenis PT. Psychiatric comorbidity in gender dysphoric adolescents. *J Child Psychol Psychiatry*. 2011;52(11):1195–1202
16. Holt V, Skagerberg E, Dunsford M. Young people with features of gender dysphoria: demographics and associated difficulties. *Clin Child Psychol Psychiatry*. 2016;21(1):108–118
17. de Graaf NM, Cohen-Kettenis PT, Carmichael P, et al. Psychological functioning in adolescents referred to specialist gender identity clinics across Europe: a clinical comparison study between four clinics. *Eur Child Adolesc Psychiatry*. 2018;27(7):909–919
18. Fisher AD, Ristori J, Castellini G, et al. Psychological characteristics of Italian gender dysphoric adolescents: a case-control study. *J Endocrinol Invest*. 2017; 40(9):953–965
19. Lindgren TW, Pauly IB. A body image scale for evaluating transsexuals. *Arch Sex Behav*. 1975;4(6):639–656
20. Birmaher B, Brent DA, Chiappetta L, Bridge J, Monga S, Baugher M. Psychometric properties of the Screen for Child Anxiety Related Emotional Disorders (SCARED): a replication study. *J Am Acad Child Adolesc Psychiatry*. 1999;38(10):1230–1236
21. Rush AJ, Trivedi MH, Ibrahim HM, et al. The 16-Item Quick Inventory of Depressive Symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. *Biol Psychiatry*. 2003;54(5):573–583
22. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 5th ed. Washington, DC: American Psychiatric Association; 2013
23. Cohen J. *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed. Mahwah, NJ: Lawrence Erlbaum Associates; 1988
24. Eisenberg ME, Gower AL, McMorris BJ, Rider GN, Shea G, Coleman E. Risk and protective factors in the lives of transgender/gender nonconforming adolescents. *J Adolesc Health*. 2017; 61(4):521–526
25. Clark TC, Lucassen MF, Bullen P, et al. The health and well-being of transgender high school students: results from the New Zealand adolescent health survey (Youth'12). *J Adolesc Health*. 2014;55(1):93–99
26. Wikipedia. Bathroom bill. Available at: https://en.wikipedia.org/wiki/Bathroom_bill. Accessed August 15, 2019
27. Aitken M, Steensma TD, Blanchard R, et al. Evidence for an altered sex ratio in clinic-referred adolescents with gender dysphoria. *J Sex Med*. 2015; 12(3):756–763

Body Dissatisfaction and Mental Health Outcomes of Youth on Gender-Affirming Hormone Therapy

Laura E. Kuper, Sunita Stewart, Stephanie Preston, May Lau and Ximena Lopez

Pediatrics 2020;145;
DOI: 10.1542/peds.2019-3006 originally published online March 27, 2020;

PEDIATRICS®
OFFICIAL JOURNAL OF THE AMERICAN ACADEMY OF PEDIATRICS

Body Dissatisfaction and Mental Health Outcomes of Youth on Gender-Affirming Hormone Therapy

Laura E. Kuper, Sunita Stewart, Stephanie Preston, May Lau and Ximena Lopez

Pediatrics 2020;145;
DOI: 10.1542/peds.2019-3006 originally published online March 27, 2020;

Summary of a recommendation by COHERE Finland 16.6.2020

Medical treatment methods for dysphoria associated with variations in gender identity in minors – recommendation

In its meeting on 11 June 2020, the Council for Choices in Health Care in Finland (COHERE Finland) adopted a recommendation on medical treatment methods for dysphoria associated with variations in the gender identity of minors

The recommendation clarifies the roles of different healthcare operators in a situation where a minor is uncertain about their gender identity. The recommendation presents the medical treatment methods that fall within the range of public healthcare services when it comes to the medical treatment of gender dysphoria in minors.
In COHERE's view, psychosocial support should be provided in school and student healthcare and in primary healthcare for the treatment of gender dysphoria due to variations in gender identity in minors, and there must be sufficient competency to provide such support. Consultation with a child or youth psychiatrist and the necessary psychiatric treatment and psychotherapy should be arranged locally according to the level of treatment needed. If a child or young person experiencing gender-related anxiety has other simultaneous psychiatric symptoms requiring specialised medical care, treatment according to the nature and severity of the disorder must be arranged within the services of their own region, as no conclusions can be drawn on the stability of gender identity during the period of disorder caused by a psychiatric illness with symptoms that hamper development.

In Finland, the diagnostics of gender identity variation, the assessment of the need for medical treatments and the planning of their implementation are centralised by law in the multi-professional research clinics of Helsinki University Central Hospital (HUS) and Tampere University Hospital (TAYS). The consultation, evaluation periods and treatments provided by the TAYS or HUS working group on the gender identity of minors shall be carried out in accordance with the following principles.

Children who have not started puberty and are experiencing persistent, severe anxiety related to gender conflict and/or identification as the other sex may be sent for a consultation visit to the research group on the gender identity of minors at TAYS or HUS. Any need for support beyond the consultation visit or need for other psychiatric treatment should be addressed by local services according to the nature and severity of the problem.

If a child is diagnosed prior to the onset of puberty with a persistent experience of identifying as the other sex and shows symptoms of gender-related anxiety, which increases in severity in puberty, the child can be guided at the onset of puberty to the research group on the gender identity of minors at TAYS or HUS for an assessment of the need for treatment to suppress puberty. Based on these assessments, puberty suppression treatment may be initiated on a case-by-case basis after careful consideration and appropriate diagnostic examinations if the medical indications for the treatment are present and there are no contraindications. Therapeutic amenorrhea, i.e. prevention of menstruation, is also medically possible.

A young person who has already undergone puberty can be sent to the research clinic on the gender identity of minors at TAYS or HUS for extensive gender identity studies if the variation in gender identity and related dysphoria do not reflect the temporary search for identity typical of the development stage of adolescence and do not subside once the young person has had the opportunity to reflect on their identity but rather their identity and personality development appear to be stable.

PO BOX 33 (Meritullinkatu 8, Helsinki)
00023 Government, Finland
www.palveluvalikoima.fi
Tel 0295 16001
e-mail:
palveluvalikoima@stm.fi

Summary of a recommendation by COHERE Finland 16.6.2020

Based on thorough, case-by-case consideration, the initiation of hormonal interventions that alter sex characteristics may be considered before the person is 18 years of age only if it can be ascertained that their identity as the other sex is of a permanent nature and causes severe dysphoria. In addition, it must be confirmed that the young person is able to understand the significance of irreversible treatments and the benefits and disadvantages associated with lifelong hormone therapy, and that no contraindications are present.

If a young person experiencing gender-related anxiety has experienced or is simultaneously experiencing psychiatric symptoms requiring specialised medical care, a gender identity assessment may be considered if the need for it continues after the other psychiatric symptoms have ceased and adolescent development is progressing normally. In this case, a young person can be sent by the specialised youth psychiatric care in their region for an extensive gender identity study by the TAYS or HUS research group on the gender identity of minors, which will begin the diagnostic studies. Based on the results of the studies, the need for and timeliness of medically justified treatments will be assessed individually.

Surgical treatments are not part of the treatment methods for dysphoria caused by gender-related conflicts in minors. The initiation and monitoring of hormonal treatments must be centralised at the research clinics on gender identity at HUS and TAYS.
Research data on the treatment of dysphoria due to gender identity conflicts in minors is limited. COHERE considers that, moving forward, multi-professional clinics specialising in the diagnostics and treatment of gender identity conflicts at HUS and TAYS should collect extensive information on the diagnostic process and the effects of different treatment methods on the mental wellbeing, social capacity and quality of life of children and youth. There is also a need for more information on the disadvantages of procedures and on people who regret them.

Link to the COHERE website: https://palveluvalikoima.fi/en/frontpage

The Council for Choices in Health Care in Finland (COHERE Finland) works in conjunction with the Ministry of Social Affairs and Health, and its task is to issue recommendations on services that should be included in the range of public health services. Further information: www.palveluvalikoima.fi.

PO BOX 33 (Meritullinkatu 8, Helsinki)
00023 Government, Finland
www.palveluvalikoima.fi
Tel 0295 16001
e-mail:
palveluvalikoima@stm.fi

Published in final edited form as:
Paediatr Drugs. 2015 August ; 17(4): 273–281. doi:10.1007/s40272-015-0130-8.

Central Precocious Puberty: Update on Diagnosis and Treatment

Melinda Chen[1] and **Erica A. Eugster**[1]

[1]Section of Pediatric Endocrinology, Department of Pediatrics, Riley Hospital for Children, Indiana University School of Medicine, 705 Riley Hospital Drive, Room # 5960, Indianapolis, IN 46202, USA

Abstract

Central precocious puberty (CPP) is characterized by the same biochemical and physical features as normally timed puberty but occurs at an abnormally early age. Most cases of CPP are seen in girls, in whom it is usually idiopathic. In contrast, ~50 % of boys with CPP have an identifiable cause. The diagnosis of CPP relies on clinical, biochemical, and radiographic features. Untreated, CPP has the potential to result in early epiphyseal fusion and a significant compromise in adult height. Thus, the main goal of therapy is preservation of height potential. The gold-standard treatment for CPP is go-nadotropin-releasing hormone (GnRH) analogs (GnRHa). Numerous preparations with a range of delivery systems and durations of action are commercially available. While the outcomes of patients treated for CPP have generally been favorable, more research about the psychological aspects, optimal monitoring, and long-term effects of all forms of GnRHa treatment is needed. Several potential therapeutic alternatives to GnRHa exist and await additional investigation.

1 Introduction

Central precocious puberty (CPP) refers to premature activation of the hypothalamic–pituitary–gonadal (HPG) axis, resulting in early development of secondary sexual characteristics. Although the exact threshold defining “normal” pubertal timing has been disputed, commonly used cutoffs to define CPP are 8 years of age for females (7.5 years for Hispanics and African Americans) and 9 years of age for males [1]. The earliest clinical manifestation of central puberty in girls is usually breast development (thelarche), followed by pubic hair (pub-arche). The pubertal growth spurt typically occurs during Tanner stage II–III, with the first menstrual period, known as menarche, usually occurring at Tanner stage IV. In boys, the initial clinical sign of central puberty is testicular enlargement and the pubertal growth spurt happens later than in girls [2, 3].

Although the precise mechanisms triggering the onset of puberty are unclear, the earliest known biochemical change during puberty is increased production of kisspeptin in the hypothalamus. While kisspeptin itself has several proposed stimulatory and inhibitory signals, which have not yet been clearly elucidated, it has been shown that increased

Correspondence to: Melinda Chen.

Conflict of interest Dr. Chen has no conflicts of interest to disclose. Dr. Eugster participates in clinical trials investigating treatment of CPP, funded by Endo Pharmaceuticals. No sources of funding were used to support the writing of this article.

kisspeptin production results in increased gonadotropin-releasing hormone (GnRH) release. Thus, a rise in kisspeptin is widely acknowledged as the seminal event that initiates HPG axis activation during puberty [2]. Inhibition of the GnRH pulse generator decreases first during sleep, resulting in an increase of nighttime luteinizing hormone (LH) pulse amplitude during early and mid-puberty. As puberty progresses, LH pulse amplitude increases during daytime hours as well, and estrogen and testosterone levels rise accordingly.

2 Etiology

CPP, for unknown reasons, is found predominantly in girls. In an observational study of the incidence of CPP in Spain, females were approximately ten times more likely to be affected than males [4], and other sources have cited a female-to-male ratio as high as 20:1 [5]. In addition, the etiology of CPP differs between the genders. While the majority of girls will have idiopathic CPP, boys are more likely to have a pathological source [1, 6]. Risk factors for CPP include a history of international adoption, as well as congenital or acquired central nervous system insults, such as hypothalamic hamartoma, septo-optic dysplasia, tumor, trauma, infection, or ischemia. Several genetic syndromes, including neurofibromatosis type 1, tuberous sclerosis, and Sturge–Weber syndrome, are associated with CPP [2]. Apart from recognized genetic syndromes, anywhere from 5.2 to 27.5 % of cases have been reported to be familial [7, 8].

Specific genetic causes of CPP have been described relatively recently. A substitution mutation in the G-protein coupled kisspeptin receptor gene *KISS1R* (formerly known as *GPR54*) was found in a patient with CPP and was associated with delayed degradation of the ligand–receptor complex within the cell membrane. This was further linked to an extended period of downstream signaling, postulated to result in increased amplitude of GnRH pulsatility [9]. An additional *KISS1R* polymorphism in the promoter region has been described in Chinese girls with CPP, though a detailed knowledge of whether or how this variant impacts the expression or function of the gene is as yet unknown [10].

A mutation in *KISS1*, encoding the ligand kisspeptin, has also been described within an amino-terminal sequence associated with protein degradation [11]. The mutated li-gand–receptor complex similarly demonstrates resistance to degradation. However, the low population frequency associated with this mutation suggests that it is a relatively uncommon cause of CPP.

More recently, ten separate heterozygous mutations in *MKRN3*, encoding makorin RING-finger protein 3, have been found in association with both sporadic and familial CPP [12–14]. *MKRN3* is a paternally expressed imprinted gene located within the region typically affected in Prader–Willi syndrome. Although the exact function of *MKRN3* in humans is as yet unknown, studies in mice have illustrated that mkrn3 mRNA is expressed in the hypothalamic arcuate nucleus, and that a decline in mkrn3 expression is temporally correlated with the rise in kiss1 expression. Other studies have postulated that down-regulation of *MKRN3* is permissive for increased GnRH pulses during puberty [13]. Thus, deficiency of this protein would be expected to result in a loss of inhibition of HPG axis activation. These mutations are thought to result in loss of function of the abnormal gene

product. In familial cases, all affected subjects have inherited mutations from their fathers. Interestingly, there was an almost equal gender distribution of CPP among affected family members [12].

Other molecular defects have been identified with less clear or weaker associations. These include single nucleotide polymorphisms (SNPs) in the *FSHB* gene and the *LHB* gene, though the resulting molecular mechanisms that cause CPP have not been identified [15]. Mutations in the Y1 subtype receptor for neuropeptide Y (*NPY*) could theoretically cause precocious puberty, as *NPY* is thought to be an inhibitor of pulsatile GnRH secretion. However, the only currently described mutation has not correlated well with an effect on function or with CPP [16]. Additional studies have investigated genes involved in hypothalamic hamartomas and have identified several with increased expression in patients with CPP [17]. *LIN28B*, which is postulated to have a role in determining the timing of pubertal development, has also been proposed as a genetic target in CPP. However, its exact role in humans is not yet clear. In addition, study findings have been contradictory, and no clinically significant mutations have yet been observed that cause a functional deficit at a molecular level [18, 19].

3 Diagnosis

3.1 Clinical Features

On initial examination of the child with CPP, bilateral testicular enlargement (≥4 cc in volume) will be apparent in males, in contrast to patients with peripheral forms of precocious puberty. Girls usually present with both breast development and pubic hair, in contrast to nonpathological entities such as premature thelarche or premature adrenarche. Other signs of pathological precocious puberty include a rapid tempo of progression and linear growth acceleration. Bone age will typically be advanced, though this is certainly not exclusive to CPP and may be seen to a milder degree in numerous other conditions [2].

3.2 Biochemical Features

A GnRH stimulation test has long been considered the gold standard for the diagnosis of CPP. However, lack of availability of synthetic GnRH in the USA has led to the use of GnRH analogs (GnRHa) for this purpose instead. While precise cutoffs are difficult to establish, a peak stimulated LH of >~8 mIU/mL after GnRH and >~5 IU/L after GnRHa are considered indicative of CPP [2, 20]. An LH/FSH [luteinizing hormone/follicle-stimulating hormone] ratio of ≥2 is also consistent with CPP. However, the results should always be interpreted in light of the specific assay performed and the available sensitivity limits. An alternative diagnostic approach has been measurement of basal ultrasensitive LH, which is typically <0.3 IU/L in prepubertal children. However, basal ultrasensitive LH is often prepubertal in early CPP and thus may be falsely reassuring [1]. Measurement of basal or stimulated sex steroids, while never sufficient alone, can be helpful in evaluation of suspected CPP. This is particularly true of testosterone, whereas random estradiol levels are often unmeasurable even when advanced pubertal development is present. Even if the laboratory evaluation is unremarkable, patients should continue to be monitored over time and retested as indicated if clinical suspicion is high [1, 2, 20].

4 Imaging

Pelvic ultrasound has been found to be a useful adjunct to support the diagnosis of CPP over other forms of puberty in girls, especially in equivocal situations. Uterine and ovarian dimensions have a stronger association with bone age than with chronological age and are correlated with CPP up to the age of 8 years [21]. While proposed cutoffs for uterine and ovarian volumes exist, these have been somewhat variable, and other studies have suggested a considerable overlap between patients with and without CPP, making reliable parameters difficult to establish. For those who present for evaluation after the age of 8 years, ultrasound parameters become even more difficult to interpret, as there is an even greater overlap in uterine and ovarian dimensions between prepubertal and early pubertal girls [21–24]. The finding of small ovarian follicles on a pelvic ultrasound is normal even in prepubertal girls [25]. Clinicians should further keep in mind that ultrasound results may be technician dependent.

The role of brain magnetic resonance imaging (MRI) in the evaluation of patients with CPP has been debated. Boys are more likely to have a pathological cause, making diagnostic imaging for intracranial pathology an essential tool in their evaluation [6]. However, controversy exists regarding recommendations in girls. When female CPP patients without neurological symptoms are screened with MRI, the incidence of positive findings is approximately 15 % [26, 27]. However, some of the abnormalities that are found may be incidental and unrelated to the CPP. In one study, 86 % of 182 girls had normal MRIs, 11 % had mild abnormalities believed to be unrelated to CPP, and 3 % had hamartomas, leading the authors to conclude that routine screening was not indicated in this population, particularly in girls older than 6 years [26]. This is in contrast to a prior study of 67 girls, in whom six of ten with MRI findings had hamartomas, while the remainder were diagnosed with an astrocytoma, teratoma, arachnoid cyst, and pineal cyst. Three of the ten had lesions requiring surgical intervention, leading the authors to conclude that MRI should be part of routine evaluation in CPP regardless of age [27]. Investigations into clinical and biochemical features of patients with intracranial pathology have suggested that younger age at onset, more rapid tempo, and higher levels of sex steroids or gonadotropins are predictive features. However, these overlap to such an extent that no specific cutoff has been identified that can be used to determine whether or not to obtain an MRI in any individual patient. For this reason, many institutions include a brain MRI as a universal part of the evaluation in all children diagnosed with CPP [26–29].

5 Treatment

The primary goal of CPP treatment is to preserve final adult height. However, it should be recognized that some patients will have a nonprogressive or slowly progressive form of CPP, and these patients can achieve normal adult height without any intervention [20]. Therefore, a period of observation is usually appropriate prior to starting treatment. In patients who do show progression of CPP, there is significant variability in the degree of height gained after discontinuation of treatment, even among patients with the same bone age [30–32]. Numerous studies have demonstrated that the greatest gain in final height is achieved in girls with onset of puberty before 6 years of age, although girls with onset between 6 and 8 years

of age may still reap some benefit from treatment. In contrast, girls aged ≥8 years have not been found to benefit from intervention in terms of height. Thus, treatment in this age group is usually not indicated. An additional issue is that outcomes of treatment are typically defined as the difference between predicted adult height at baseline and the actual height achieved. Unfortunately, height prediction methods are notoriously flawed [33] and have often been found to overpredict height in the setting of early puberty. Therefore, it is impossible to predict the precise amount of additional height that will be gained by an individual patient as a result of putting puberty on hold. While preliminary evidence suggests that electronic methods of bone age assessment may be more accurate, there is minimal information available thus far about their use in precocious puberty [34]. Evidence regarding treatment benefit in males is more limited, as they comprise a relatively small proportion of patients with CPP. The existing data suggest a significant improvement in final height after treatment of CPP in boys [35], though the same measurement and prediction limitations exist.

Concerns about psychosocial functioning are often used as a justification for treatment of CPP. However, the existing data regarding the psychological aspects of CPP are limited and inconsistent. Insufficient controls and methodological problems render many studies difficult to interpret, compounded by the use of several different assessments, which make comparisons difficult. The current data do not consistently support problems in regard to body image, self-esteem, or sexual behavior in patients with CPP. Differences, where found, tend to be modest and suggest that patients with CPP may engage in psychosexual behaviors only slightly earlier than children with on-time puberty. The prevalence of psychopathology does not seem to differ from that in the general population [36]. Similarly, one study of girls with CPP and their mothers at the time of diagnosis found no difference in psychological distress as compared with girls who had early normal puberty, even prior to treatment [37]. At this point, there is no consensus regarding whether CPP is associated with psychological distress and/or whether treatment ameliorates these problems, and more data in this area are needed [20].

GnRHa are well established as a standard of care for the treatment of CPP worldwide. While numerous delivery systems and routes of administration exist, depot intramuscular injections or sustained-release preparations have been most widely used. These drugs are believed to work by providing a steady concentration of GnRH activity instead of the pulsatile variation in levels characteristic of native GnRH release, which results in paradoxical down-regulation and suppression of the HPG axis. Monthly depot leuprolide acetate has been the most common form of injection therapy used in the USA. Although extended-release 3-monthly depot leuprolide preparations have been available in Europe and elsewhere for many years, they have been approved by the US Food and Drug Administration (FDA) only recently and are available in 11.25 and 30 mg dosage forms. Although patients on 11.25 mg 3-monthly injections have consistently been shown to have higher stimulated LH and FSH levels than patients receiving 7.5 mg monthly injections or 22.5 mg 3-monthly injections, this has not been accompanied by significant differences in sex steroid levels or clinical parameters [38, 39]. Additional information about these preparations has been derived from a phase III open-label study involving patients receiving 3-monthly depot leuprolide acetate at 11.25 and 30 mg doses for 36 months [40]. Of 72 patients, only two discontinued therapy prior to 36

months because of treatment failure, while 20 discontinued therapy to undergo age-appropriate puberty and 24 continued to receive 3-monthly depot leuprolide for the full study period. As in previous studies, LH escape was seen in a minority of patients on stimulation testing, but this did not correlate with clinical features suggesting lack of suppression. Thus, 3-monthly depot leuprolide seems to be both safe and effective for long-term use [38–40].

Adverse effects are similar for 1- and 3-monthly depot injections and include local reactions and pain at the injection site. Sterile abscess formation has been reported after administration of long-acting injection formulations. Although children who experience sterile abscess formation from long-acting preparations have subsequently been treated successfully with daily leuprolide, there are reported cases in the adult literature of resistance to GnRHa following sterile abscesses [41].

A popular alternative approach to depot GnRHa injections is the histrelin implant, which was approved by the FDA in 2007. This nonbiodegradable implant is made of a flexible hydrogel containing 50 mg of the potent GnRHa histrelin and is placed subcutaneously, usually in the inner aspect of the upper arm. The initial histrelin implant was first developed for treatment of metastatic prostate cancer, where it was found to successfully suppress LH and testosterone levels for up to 1 year. The implant was later reformulated to release histrelin at a higher dose of 65 lg/day for use in children with CPP. An initial pilot study in 11 girls previously treated with depot triptorelin showed satisfactory maintenance of LH and FSH suppression on stimulation tests. This was accompanied by clinical evidence of pubertal suppression, including regression of breast development, a decrease in growth velocity, and a decline in bone age advancement over 15 months. In addition to satisfactory clinical benefit, parents reported less discomfort and lifestyle interference overall than with monthly injections [42]. Following this initial report, a phase III study in 36 patients with CPP demonstrated profound suppression of the HPG axis within 1 month of implantation whether subjects were naïve or previously treated with a GnRHa [43]. The long-term extension phase of this study has now been completed and demonstrated significant improvements in predicted adult height after up to 6 years of sequential annual histrelin implants [44]. Reassuringly, body mass index (BMI) z-scores remained normal throughout the treatment interval.

A significant refinement of the histrelin implant as a therapeutic option has been the recognition that a single implant lasts at least 2 years. Given the known rate of release of 65 mcg of histrelin per day, a 50 mg implant should theoretically last 2 years. That this is indeed the case was demonstrated in a prospective study in 33 children with CPP in whom a single implant was left in place for 2 years. Peak stimulated LH levels at 12 and 24 months were equivalent, and clinical indices of CPP improved progressively. Use of a single implant for 2 years has the potential to significantly decrease costs and numbers of surgical procedures in children treated with this modality [45].

The most common adverse event associated with the histrelin implant is breakage and/or difficulty with localization of the device. These events occur only during explantation and have been noted to take place in 15–39 % of procedures, with a higher likelihood of

breakage when the implant is left in place for 2 years [44–47]. Additional reported adverse events include local reactions, which are, for the most part, minor and self-limited. Sterile abscess formation [41], keloids [44], site infection [45], and implant extrusion [42] have rarely been reported. Placement and removal of the implant requires a minor surgical procedure. This is typically accomplished in an outpatient setting, using local anesthesia with the addition of conscious sedation if necessary [47]. In rare cases of difficulty with implant localization, ultrasound has proved to be a useful modality. Characteristics of the most frequently used GnRHa are summarized in Table 1.

6 Adjunctive Treatments

6.1 Nonaromatizable Anabolic Steroids

Oxandrolone has been used to improve growth in patients for other indications. The exact mechanism has not been elucidated, but a stimulatory effect on the growth plate has been postulated. A small nonrandomized study of ten patients receiving oxandrolone in addition to GnRHa for severe deceleration of growth velocity during treatment for CPP suggested that this combination might improve adult height, compared with GnRHa alone. However, larger randomized studies have not been performed [48].

6.2 Growth Hormone

Small and nonrandomized studies have demonstrated a significant improvement in final adult height over pre-treated predicted adult height in patients treated with GnRHa and growth hormone (GH) as compared with patients treated with GnRHa alone. However, larger randomized studies are currently lacking, and routine use of GH in this setting is not recommended [49].

6.3 Aromatase Inhibitors

Aromatase inhibitors have the potential to attenuate estrogenic effects on skeletal maturation and to delay epiphyseal fusion. A small randomized study suggested slower bone age advancement and improved adult height in Chinese boys with CPP receiving letrozole [50]. However, in general, the experience with these compounds in CPP has been very limited.

7 Monitoring

Children who are being treated for CPP should receive regular follow-up during which pubertal progression or suppression can be followed and documented. Tanner staging, determination of growth velocity, and assessment of skeletal maturation via bone age radiographs are all important indices of suppression. Whether laboratory testing should be routinely included during follow-up is controversial. While several different strategies for biochemical testing exist, no gold standard for how best to monitor children undergoing treatment for CPP has been established. A GnRH- or leuprolide-stimulated peak LH should be <4 IU/L in adequately suppressed children, and random serum gonadotropin levels should theoretically be in the prepubertal range (ultrasensitive LH <0.3 IU/L). However, random ultrasensitive LH levels have been noted to be elevated above prepubertal levels in children who are well suppressed on GnRHa therapy across all forms of treatment, including

the histrelin implant [51, 52]. Therefore, the utility of measuring random LH levels in children undergoing treatment for CPP is highly questionable.

8 Resumption of the Native Hypothalamic–Pituitary–Gonadal Axis

In studies following patients beyond discontinuation of treatment, the mean time from cessation of injectable depot GnRHa to menses has been found to be 1.5 ± 0.5 years. Some studies have found a slightly shorter time to menses in girls who experienced menarche before treatment than in those who did not. Although less is known about boys, the existing data suggest that clinicians can expect advancement of the Tanner stage within 6 months of discontinuation of treatment [52].

Because use of the histrelin implant is more recent, the data are somewhat more limited but thus far seem to indicate similar results, with the average time from explantation to menarche being 12.75 (95 % confidence interval 9.6–15.9) months, with a range of 2–36 months. Likewise, in males, resumption of pubertal progression was seen on examination within 1 year. A negative trend has been noted between the total duration of GnRHa therapy and the time to menarche, whereas the age at explantation and the time to menarche were significantly inversely correlated [53].

9 Outcomes

Girls with CPP have been found to have a higher BMI than their peers at diagnosis. However, this observation is confounded by the natural increase in BMI during puberty. Indeed, some authors have found that while BMI increases in general during treatment, the overall BMI standard deviation score (SDS) does not change. When BMI is evaluated after GnRHa treatment has been completed, there does not appear to be an adverse effect of treatment on BMI in girls with CPP, nor does there appear to be a large impact of CPP itself on BMI at adult height [31, 54, 55].

Results regarding the incidence of polycystic ovary syndrome (PCOS) in GnRHa-treated patients have been quite variable and contradictory, with some authors finding markedly increased rates of PCOS and other authors finding little or no difference. These results are even more difficult to interpret, as multiple criteria for the diagnosis of PCOS exist. Currently, no consensus exists on whether CPP or treatment with GnRHa results in an increased risk of PCOS [2, 54, 56].

Although bone mineral density (BMD) has been seen to be slightly reduced during treatment in girls, these changes do not appear to be sustained. This decrease is thought to be secondary to suppression of ovarian function. However, after treatment is discontinued and ovarian activity resumes, BMD is regained, and so girls are not significantly different from their peers without CPP, according to evaluation at adult height [31].

10 Reproductive Function and Fertility

Limited information exists regarding the long-term effects of treatment for CPP on endocrine and reproductive function. In one study of 49 females receiving monthly depot

leuprolide, 20 were followed to adulthood (age 18–26 years). Of these, 80 % reported regular menstrual cycles. Seven of 20 women reported a total of 12 pregnancies, with six live births, five spontaneous or elective terminations, one ongoing pregnancy, and no reports of stillbirth [52]. Though achievement of short-term treatment goals and resumption of puberty seem to be similar in girls treated with 3-monthly leuprolide and the histrelin implant, it remains to be seen whether similar long-term results can be expected. In addition, long-term data are notably lacking for all forms of treatment with regard to fertility and endocrine function beyond the third decade, as well as the timing of menopause.

11 Future Directions

Although multiple preparations now exist for GnRHa treatment of CPP, further options are under investigation or may be considered. A 6-month formulation of triptorelin, for example, is currently under investigation, providing the potential for even less frequent dosing for those who do not wish to undergo a procedure for the histrelin implant [57].

Other targets for therapy could also be considered. Because GnRH agonists work by stimulation of receptors, leading to desensitization, there is an initial period of increased stimulation, leading to an LH flare, which sometimes precipitates vaginal bleeding in girls with advanced pubertal development before suppression takes place. A GnRH antagonist, however, would theoretically forego this initial phase by disrupting LH pulsatility without an initial flare. Kisspeptin agonists and antagonists, by acting upstream of GnRH, would be expected to have effects similar to those of GnRH agonist and antagonist therapies, respectively. However, since kisspeptin analogs would not work directly at the gonadotropin receptor, they would have the additional theoretical benefit of interrupting pulsatile GnRH and gonadotropin secretion without lowering gonadotropin release below basal secretory levels. Therefore, sex steroid levels under treatment with these agents could be expected to more closely mimic normal physiology [58]. Though use of kisspeptin antagonists in humans has not yet been studied, animal studies have suggested that kisspeptin analogs are able to cross the blood–brain barrier and suppress puberty [58].

Because existing biochemical markers can be unreliable, monitoring of treatment is also a worthwhile area of research. Markers under investigation include free alpha-subunit (FAS), which rises with suppression of the HPG axis. Though GnRHa levels decrease gradually with discontinuation of depot intramuscular injections, FAS levels are seen to acutely decrease within days of histrelin implant removal, preceding LH, FSH, and estradiol rises by weeks. In one case, elevated FAS levels beyond the expected time period were attributable to retained histrelin implant fragments and fell only after all fragments had been removed, highlighting the utility of FAS as a target for relatively rapid assessment of HPG axis recovery. A short “rebound” elevation in FAS can be seen in patients 3–8 weeks after histrelin implant removal. This effect is short lived and self limited, although the reasons for it are unclear [59].

12 Conclusion

CPP is seen most often in girls and is associated with a multitude of conditions. A substantial proportion (over a quarter) of cases are familial, and genetic causes have begun to be elucidated. The diagnosis is based on a combination of clinical and biochemical factors. Treatment with a GnRHa provides the greatest potential benefit for patients who are younger at the time of onset of CPP. Multiple treatment options are available, and more recent options have the benefit of less frequent dosing, with potential for improved compliance. Though several adjunctive treatments have been proposed, evidence supporting these treatments is generally sparse in CPP, and thus they cannot be recommended for routine use. Biochemical markers, bone age, and growth velocity should be followed during treatment to ensure efficacy. The available evidence shows that GnRHa are safe and effective, and long-term data suggest that reproductive function is satisfactory after discontinuation of treatment. However, long-term data, particularly regarding the newer formulations, are still lacking. Continued pharmacological and molecular genetic investigation and rigorously conducted prospective studies will continue to enhance knowledge and optimize treatment in children with CPP.

Acknowledgments

None.

References

1. Nebesio TD, Eugster EA. Current concepts in normal and abnormal puberty. Curr Probl Pediatr Adolesc Health Care. 2007; 37(2):50–72. DOI: 10.1016/j.cppeds.2006.10.005 [PubMed: 17223057]
2. Fuqua JS. Treatment and outcomes of precocious puberty: an update. J Clin Endocrinol Metab. 2013; 98(6):2198–207. DOI: 10.1210/jc.2013-1024 [PubMed: 23515450]
3. Tanner JM, Davies PSW. Clinical longitudinal standards for height and height velocity for North American children. J Pediatr. 1985; 107(3):317–29. DOI: 10.1016/s0022-3476(85)80501-1 [PubMed: 3875704]
4. Soriano-Guillen L, Corripio R, Labarta JI, Canete R, Castro-Feijoo L, Espino R, et al. Central precocious puberty in children living in Spain: incidence, prevalence, and influence of adoption and immigration. J Clin Endocrinol Metab. 2010; 95(9):4305–13. DOI: 10.1210/jc.2010-1025 [PubMed: 20554707]
5. Lee PA, Neely EK, Fuqua J, Yang D, Larsen LM, Mattia-Gold-berg C, et al. Efficacy of leuprolide acetate 1-month depot for central precocious puberty (CPP): growth outcomes during a prospective, longitudinal study. Int J Pediatr Endocrinol. 2011; 2011(1):7.doi: 10.1186/1687-9856-2011-7 [PubMed: 21860633]
6. Choi KH, Chung SJ, Kang MJ, Yoon JY, Lee JE, Lee YA, et al. Boys with precocious or early puberty: incidence of pathological brain magnetic resonance imaging findings and factors related to newly developed brain lesions. Ann Pediatr Endocrinol Metab. 2013; 18(4):183–90. DOI: 10.6065/apem.2013.18.4.183 [PubMed: 24904875]
7. de Vries L, Kauschansky A, Shohat M, Phillip M. Familial central precocious puberty suggests autosomal dominant inheritance. J Clin Endocrinol Metab. 2004; 89(4):1794–800. DOI: 10.1210/jc.2003-030361 [PubMed: 15070947]
8. Rohn R, Rousonelos G. Familial sexual precocity. Am J Dis Child. 1986; 140(8):741–2. DOI: 10.1001/archpedi.1986.02140220023017
9. Teles MG, Bianco SD, Brito VN, Trarbach EB, Kuohung W, Xu S, et al. A GPR54-activating mutation in a patient with central precocious puberty. N Engl J Med. 2008; 358(7):709–15. DOI: 10.1056/NEJMoa073443 [PubMed: 18272894]

10. Luan X, Yu H, Wei X, Zhou Y, Wang W, Li P, et al. GPR54 polymorphisms in Chinese girls with central precocious puberty. Neuroendocrinology. 2007; 86(2):77–83. DOI: 10.1159/000107511 [PubMed: 17700012]
11. Silveira LG, Noel SD, Silveira-Neto AP, Abreu AP, Brito VN, Santos MG, et al. Mutations of the KISS1 gene in disorders of puberty. J Clin Endocrinol Metab. 2010; 95(5):2276–80. DOI: 10.1210/jc.2009-2421 [PubMed: 20237166]
12. Abreu AP, Dauber A, Macedo DB, Noel SD, Brito VN, Gill JC, et al. Central precocious puberty caused by mutations in the imprinted gene MKRN3. N Engl J Med. 2013; 368(26):2467–75. DOI: 10.1056/NEJMoa1302160 [PubMed: 23738509]
13. Macedo DB, Abreu AP, Reis AC, Montenegro LR, Dauber A, Beneduzzi D, et al. Central precocious puberty that appears to be sporadic caused by paternally inherited mutations in the imprinted gene makorin ring finger 3. J Clin Endocrinol Metab. 2014; 99(6):E1097–103. DOI: 10.1210/jc.2013-3126 [PubMed: 24628548]
14. Settas N, Dacou-Voutetakis C, Karantza M, Kanaka-Gantenbein C, Chrousos GP, Voutetakis A. Central precocious puberty in a girl and early puberty in her brother caused by a novel mutation in the MKRN3 gene. J Clin Endocrinol Metab. 2014; 99(4):E647–51. DOI: 10.1210/jc.2013-4084 [PubMed: 24438377]
15. Zhao Y, Chen T, Zhou Y, Li K, Xiao J. An association study between the genetic polymorphisms within GnRHI, LHbeta and FSHbeta genes and central precocious puberty in Chinese girls. Neurosci Lett. 2010; 486(3):188–92. DOI: 10.1016/j.neulet.2010.09.049 [PubMed: 20869425]
16. Freitas KC, Ryan G, Brito VN, Tao YX, Costa EM, Mendonca BB, et al. Molecular analysis of the neuropeptide Y1 receptor gene in human idiopathic gonadotropin-dependent precocious puberty and isolated hypogonadotropic hypogonadism. Fertil Steril. 2007; 87(3):627–34. DOI: 10.1016/j.fertnstert.2006.07.1519 [PubMed: 17140570]
17. Parent AS, Matagne V, Westphal M, Heger S, Ojeda S, Jung H. Gene expression profiling of hypothalamic hamartomas: a search for genes associated with central precocious puberty. Horm Res. 2008; 69(2):114–23. DOI: 10.1159/000111815 [PubMed: 18059092]
18. Park SW, Lee ST, Sohn YB, Cho SY, Kim SH, Kim SJ, et al. LIN28B polymorphisms are associated with central precocious puberty and early puberty in girls. Korean J Pediatr. 2012; 55(10):388–92. DOI: 10.3345/kjp.2012.55.10.388 [PubMed: 23133486]
19. Silveira-Neto AP, Leal LF, Emerman AB, Henderson KD, Piskounova E, Henderson BE, et al. Absence of functional LIN28B mutations in a large cohort of patients with idiopathic central precocious puberty. Horm Res Paediatr. 2012; 78(3):144–50. DOI: 10.1159/000342212 [PubMed: 22964795]
20. Carel JC, Eugster EA, Rogol A, Ghizzoni L, Palmert MR, et al. Group E-LGACC. Consensus statement on the use of gonado-tropin-releasing hormone analogs in children. Pediatrics. 2009; 123(4):e752–62. DOI: 10.1542/peds.2008-1783 [PubMed: 19332438]
21. Eksioglu AS, Yilmaz S, Cetinkaya S, Cinar G, Yildiz YT, Aycan Z. Value of pelvic sonography in the diagnosis of various forms of precocious puberty in girls. J Clin Ultrasound. 2013; 41(2):84–93. DOI: 10.1002/jcu.22004 [PubMed: 23124596]
22. Badouraki M, Christoforidis A, Economou I, Dimitriadis AS, Katzos G. Evaluation of pelvic ultrasonography in the diagnosis and differentiation of various forms of sexual precocity in girls. Ultrasound Obstetr Gynecol. 2008; 32(6):819–27. DOI: 10.1002/uog.6148
23. de Vries L, Horev G, Schwartz M, Phillip M. Ultrasonographic and clinical parameters for early differentiation between precocious puberty and premature thelarche. Eur J Endocrinol. 2006; 154(6):891–8. DOI: 10.1530/eje.1.02151 [PubMed: 16728550]
24. Herter LD, Golendziner E, Flores JAM, Moretto M, Di Domenico K, Becker E, et al. Ovarian and uterine findings in pelvic sonography: comparison between prepubertal girls, girls with isolated thelarche, and girls with central precocious puberty. J Ultrasound Med. 2002; 21(11):1237–46. [PubMed: 12418765]
25. Pienkowski C, Cartault A, Carfagna L, Ernoult P, Vial J, Lemasson F, et al. Ovarian cysts in prepubertal girls. Endocr Dev. 2012; 22:101–11. DOI: 10.1159/000326627 [PubMed: 22846524]
26. Pedicelli S, Alessio P, Scire G, Cappa M, Cianfarani S. Routine screening by brain magnetic resonance imaging is not indicated in every girl with onset of puberty between the ages of 6 and 8

years. J Clin Endocrinol Metab. 2014; 99(12):4455–61. DOI: 10.1210/jc.2014-2702 [PubMed: 25238205]

27. Ng SM. Cranial MRI scans are indicated in all girls with central precocious puberty. Arch Dis Child. 2003; 88(5):414–8. DOI: 10.1136/adc.88.5.414 [PubMed: 12716713]
28. Chalumeau M, Hadjiathanasiou CG, Ng SM, Cassio A, Mul D, Cisternino M, et al. Selecting girls with precocious puberty for brain imaging: validation of European evidence-based diagnosis rule. J Pediatr. 2003; 143(4):445–50. DOI: 10.1067/s0022-3476(03)00328-7 [PubMed: 14571217]
29. Mogensen SS, Aksglaede L, Mouritsen A, Sorensen K, Main KM, Gideon P, et al. Pathological and incidental findings on brain MRI in a single-center study of 229 consecutive girls with early or precocious puberty. PloS One. 2012; 7(1):e29829.doi: 10.1371/journal.pone.0029829 [PubMed: 22253792]
30. Carel JC, Roger M, Ispas S, Tondu F, Lahlou N, Blumberg J, et al. Final height after long-term treatment with triptorelin slow release for central precocious puberty: importance of statural growth after interruption of treatment. French Study Group of Decapeptyl in Precocious Puberty. J Clin Endocrinol Metab. 1999; 84(6):1973–8. DOI: 10.1210/jcem.84.6.5647 [PubMed: 10372696]
31. Pasquino AM, Pucarelli I, Accardo F, Demiraj V, Segni M, Di Nardo R. Long-term observation of 87 girls with idiopathic central precocious puberty treated with gonadotropin-releasing hormone analogs: impact on adult height, body mass index, bone mineral content, and reproductive function. J Clin Endocrinol Metab. 2008; 93(1):190–5. DOI: 10.1210/jc.2007-1216 [PubMed: 17940112]
32. Lazar L, Padoa A, Phillip M. Growth pattern and final height after cessation of gonadotropin-suppressive therapy in girls with central sexual precocity. J Clin Endocrinol Metab. 2007; 92(9): 3483–9. DOI: 10.1210/jc.2007-0321 [PubMed: 17579199]
33. Carel JC, Lahlou N, Roger M, Chaussain JL. Precocious puberty and statural growth. Hum Reprod Update. 2004; 10(2):135–47. DOI: 10.1093/humupd/dmh012 [PubMed: 15073143]
34. Thodberg HH. Clinical review: an automated method for determination of bone age. J Clin Endocrinol Metab. 2009; 94(7):2239–44. DOI: 10.1210/jc.2008-2474 [PubMed: 19401365]
35. Mul D, Bertelloni S, Carel JC, Saggese G, Chaussain JL, Oostdijk W. Effect of gonadotropin-releasing hormone agonist treatment in boys with central precocious puberty: final height results. Horm Res. 2002; 58(1):1–7. DOI: 10.1159/000063209
36. Walvoord EC, Mazur T. Behavioral problems and idiopathic central precocious puberty: fact or fiction? Pediatr Endocrinol Rev. 2007; 4(S3):306–12.
37. Schoelwer MJ, Donahue KL, Bryk K, Didrick P, Berenbaum SA, Eugster EA. Psychological assessment of mothers and their daughters at the time of diagnosis of precocious puberty. Int J Pediatr Endocrinol. 2015; 2015(1):5.doi: 10.1186/s13633-015-0001-7 [PubMed: 25780366]
38. Badaru A, Wilson DM, Bachrach LK, Fechner P, Gandrud LM, Durham E, et al. Sequential comparisons of one-month and three-month depot leuprolide regimens in central precocious puberty. J Clin Endocrinol Metab. 2006; 91(5):1862–7. DOI: 10.1210/jc.2005-1500 [PubMed: 16449344]
39. Fuld K, Chi C, Neely EK. A randomized trial of 1- and 3-month depot leuprolide doses in the treatment of central precocious puberty. J Pediatr. 2011; 159(6):982–7.e1. doi:10.1016/j.jpeds.2011.05.036. [PubMed: 21798557]
40. Lee PA, Klein K, Mauras N, Lev-Vaisler T, Bacher P. 36-month treatment experience of two doses of leuprolide acetate 3-month depot for children with central precocious puberty. J Clin Endocrinol Metab. 2014; 99(9):3153–9. DOI: 10.1210/jc.2013-4471
41. Miller BS, Shukla AR. Sterile abscess formation in response to two separate branded long-acting gonadotropin-releasing hormone agonists. Clin Ther. 2010; 32(10):1749–51. DOI: 10.1016/j.clinthera.2010.09.009 [PubMed: 21194598]
42. Hirsch HJ, Gillis D, Strich D, Chertin B, Farkas A, Lindenberg T, et al. The histrelin implant: a novel treatment for central precocious puberty. Pediatrics. 2005; 116(6):e798–802. DOI: 10.1542/peds.2005-0538 [PubMed: 16322137]
43. Eugster EA, Clarke W, Kletter GB, Lee PA, Neely EK, Reiter EO, et al. Efficacy and safety of histrelin subdermal implant in children with central precocious puberty: a multicenter trial. J Clin En-docrinol Metab. 2007; 92(5):1697–704. DOI: 10.1210/jc.2006-2479

44. Silverman, LA., Neely, EK., Kletter, GB., Lewis, K., Chitra, S., Terleckyj, O., Eugster, EA. J Clin Endocrinol Metab. Long-term continuous suppression with once-yearly histrelin subcutaneous implants for the treatment of central precocious puberty: a final report of a phase 3 multicenter trial. Epub 2015 Mar 24
45. Lewis KA, Goldyn AK, West KW, Eugster EA. A single histrelin implant is effective for 2 years for treatment of central precocious puberty. J Pediatr. 2013; 163(4):1214–6. DOI: 10.1016/j.jpeds.2013.05.033 [PubMed: 23809043]
46. Rahhal S, Clarke WL, Kletter GB, Lee PA, Neely EK, Reiter EO, et al. Results of a second year of therapy with the 12-month histrelin implant for the treatment of central precocious puberty. Int J Pediatr Endocrinol. 2009; 2009:812517.doi: 10.1155/2009/812517 [PubMed: 19956699]
47. Davis JS, Alkhoury F, Burnweit C. Surgical and anesthetic considerations in histrelin capsule implantation for the treatment of precocious puberty. J Pediatr Surg. 2014; 49(5):807–10. DOI: 10.1016/j.jpedsurg.2014.02.067 [PubMed: 24851775]
48. Vottero A, Pedori S, Verna M, Pagano B, Cappa M, Loche S, et al. Final height in girls with central idiopathic precocious puberty treated with gonadotropin-releasing hormone analog and oxandrolone. J Clin Endocrinol Metab. 2006; 91(4):1284–7. DOI: 10.1210/jc.2005-1693 [PubMed: 16449342]
49. Pucarelli I, Segni M, Ortore M, Arcadi E, Pasquino AM. Effects of combined gonadotropin-releasing hormone agonist and growth hormone therapy on adult height in precocious puberty: a further contribution. J Pediatr Endocrinol Metab. 2003; 16(7):1005–10. [PubMed: 14513877]
50. Zhao X, Zhang Q. Clinical efficacy of letrozole in boys with idiopathic central precocious puberty. Chin J Contemp Pediatr. 2014; 16(4):397–400. DOI: 10.7499/j.issn.1008-8830.2014.04.018
51. Lewis KA, Eugster EA. Random luteinizing hormone often remains pubertal in children treated with the histrelin implant for central precocious puberty. J Pediatr. 2013; 162(3):562–5. DOI: 10.1016/j.jpeds.2012.08.038 [PubMed: 23040793]
52. Neely EK, Lee PA, Bloch CA, Larsen L, Yang D, Mattia-Goldberg C, et al. Leuprolide acetate 1-month depot for central precocious puberty: hormonal suppression and recovery. Int J Pediatr Endocrinol. 2010; 2010:398639.doi: 10.1155/2010/398639 [PubMed: 21437000]
53. Fisher MM, Lemay D, Eugster EA. Resumption of puberty in girls and boys following removal of the histrelin implant. J Pediatr. 2014; 164(4):912–6.e1. DOI: 10.1016/j.jpeds.2013.12.009 [PubMed: 24433825]
54. Magiakou MA, Manousaki D, Papadaki M, Hadjidakis D, Le-vidou G, Vakaki M, et al. The efficacy and safety of go-nadotropin-releasing hormone analog treatment in childhood and adolescence: a single center, long-term follow-up study. J Clin Endocrinol Metab. 2010; 95(1): 109–17. DOI: 10.1210/jc.2009-0793 [PubMed: 19897682]
55. Poomthavorn P, Suphasit R, Mahachoklertwattana P. Adult height, body mass index and time of menarche of girls with id-iopathic central precocious puberty after gonadotropin-releasing hormone analogue treatment. Gynecol Endocrinol. 2011; 27(8):524–8. DOI: 10.3109/09513590.2010.507289 [PubMed: 21501002]
56. Franceschi R, Gaudino R, Marcolongo A, Gallo MC, Rossi L, Antoniazzi F, et al. Prevalence of polycystic ovary syndrome in young women who had idiopathic central precocious puberty. Fertil Steril. 2010; 93(4):1185–91. DOI: 10.1016/j.fertnstert.2008.11.016 [PubMed: 19135667]
57. Dajani, T., Reiner, B., Salem, G., Shea, H., Rappaport, M., Alzohaili, O., Van Meter, Q., Domek, D., Bethin, K., Kaplowitz, P., Klein, K., Merritt, D., Rose, S., Kletter, G., Aisenberg, J., Brenner, D., Rogers, D., Silverman, L., Lee, P., Gomez, R., Cassorla, F., Yang, J., Eugster, E., Flores, O., Wright, N. ClinicalTrials.gov [Internet]. National Library of Medicine (US), Bethesda (MD): 2014. Efficacy, safety, and pharmacokinetics (PK) of triptorelin 6-month formulation in patients with central precocious puberty. https://clinicaltrials.gov/ct2/show/NCT01467882
58. Pinilla L, Aguilar E, Dieguez C, Millar RP, Tena-Sempere M. Kisspeptins and reproduction: physiological roles and regulatory mechanisms. Physiol Rev. 2012; 92(3):1235–316. DOI: 10.1152/physrev.00037.2010 [PubMed: 22811428]
59. Hirsch HJ, Lahlou N, Gillis D, Strich D, Rosenberg-Hagen B, Chertin B, et al. Free alpha-subunit is the most sensitive marker of gonadotropin recovery after treatment of central precocious puberty with the histrelin implant. J Clin Endocrinol Metab. 2010; 95(6):2841–4. DOI: 10.1210/jc.2009-2078 [PubMed: 20339028]

Key Points

Molecular genetic etiologies of central precocious puberty (CPP) are beginning to be elucidated.

Evaluation of CPP should be based on a combination of clinical and biochemical factors, as each parameter has specific limitations.

The gold-standard treatment for CPP is gonadotropin-releasing hormone (GnRH) analogs (GnRHa).

GnRHa provide sustained high concentrations of GnRH, resulting in downregulation of the hypothalamic–pituitary–gonadal axis.

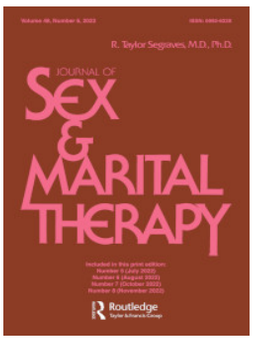
Multiple formulations of GnRHa exist. Although minor differences in gonadotropin levels are observed, all available GnRHa appear to be equally effective in terms of clinical parameters.

Long-term outcomes of children treated with GnRHa for CPP are reassuring with regard to fertility, body mass index, and bone mineral density.

Table 1

Characteristics of the most frequently prescribed gonadotropin-releasing hormone analogs for the treatment of central precocious puberty

Treatment	Administration	Frequency	Dose	Adverse effects	Advantages	Disadvantages
Leuprolide acetate	Intramuscular	Monthly	7.5 mg 11.25 mg 15 mg	Local reactions; pain at injection site; sterile abscess		Frequent injections
		3-monthly	11.25 mg 30 mg		Less frequent dosing	
Triptorelin (available only in Europe)	Intramuscular	Monthly 3-monthly	3.75 mg 11.25 mg	Local reactions; pain at injection site; sterile abscess	Less frequent dosing	Frequent injections
Histrelin implant	Subcutaneous	1–2 years	50 mg implants (65 mcg/day)	Discomfort at implant site; local reactions; sterile abscess; implant extrusion	Infrequent replacement; minimal opportunity for noncompliance	Conscious or general sedation may be required; implant breakage makes removal difficult



Routledge
Taylor & Francis Group

Journal of Sex & Marital Therapy

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/usmt20

The Dutch Protocol for Juvenile Transsexuals: Origins and Evidence

Michael Biggs

To cite this article: Michael Biggs (2022): The Dutch Protocol for Juvenile Transsexuals: Origins and Evidence, Journal of Sex & Marital Therapy, DOI: 10.1080/0092623X.2022.2121238

To link to this article: https://doi.org/10.1080/0092623X.2022.2121238

Published online: 19 Sep 2022.

JOURNAL OF SEX & MARITAL THERAPY
https://doi.org/10.1080/0092623X.2022.2121238

The Dutch Protocol for Juvenile Transsexuals: Origins and Evidence

Michael Biggs

Department of Sociology, University of Oxford, Oxford, UK

ABSTRACT

It has been a quarter of a century since Dutch clinicians proposed puberty suppression as an intervention for "juvenile transsexuals," which became the international standard for treating gender dysphoria. This paper reviews the history of this intervention and scrutinizes the evidence adduced to support it. The intervention was justified by claims that it was reversible and that it was a tool for diagnosis, but these claims are increasingly implausible. The main evidence for the Dutch protocol came from a longitudinal study of 70 adolescents who had been subjected to puberty suppression followed by cross-sex hormones and surgery. Their outcomes shortly after surgery appeared positive, except for the one patient who died, but these findings rested on a small number of observations and incommensurable measures of gender dysphoria. A replication study conducted in Britain found no improvement. While some effects of puberty suppression have been carefully studied, such as on bone density, others have been ignored, like on sexual functioning.

The use of Gonadotropin-Releasing Hormone agonist (GnRHa) drugs to suppress puberty in "juvenile transsexuals" was first proposed in print in the mid-1990s (Gooren & Delemarre-van de Waal, 1996). Developed by three clinicians at Utrecht and Amsterdam, this intervention became known as the Dutch protocol. It rapidly became standard practice in the treatment of adolescents diagnosed with gender dysphoria (HBIGDA, 2001). This intervention has been described in several manifestos by its proponents (e.g. de Vries & Cohen-Kettenis, 2012; Delemarre-van de Waal, 2014; Delemarre-van de Waal & Cohen-Kettenis, 2006) and subjected to brief critical commentaries (Byng et al., 2018; Laidlaw et al., 2019; Levine et al., 2022). The aim of this paper to provide an historical account of the invention of the Dutch protocol and a critical review of the evidence that has accumulated in the quarter of a century since it was proposed.

Before proceeding, some definitions are in order. Gender dysphoria will be used here to describe a persistent desire to become the opposite sex (Zucker, 2010). Medical terminology has changed over time, from "gender identity disorder" and "transsexualism" (both introduced in the *Diagnostic and Statistical Manual of Mental Disorders–III* in 1980) to "gender dysphoria" (as renamed in the 2013 *DSM–5*) and "gender incongruence" (as renamed in the 2019 *International Classification of Diseases–11*). There is no need to dwell on these diagnostic criteria because the condition in practice is defined by the patient's wish for endocrinological and surgical interventions. In the nomenclature of transgender medicine, "puberty blockers" denote GnRHa drugs (alternatively known as Luteinizing Hormone-Releasing Hormone agonists) which stop the production of sex hormones.[1] Drugs in this class include triptorelin (branded Decapeptyl or

CONTACT Michael Biggs michael.biggs@sociology.ox.ac.uk Department of Sociology, University of Oxford, UK

Gonapeptyl), which is used in the Netherlands and Britain, and leuprorelin (branded Lupron) in North America. GnRHa drugs are licensed to treat several medical conditions including precocious puberty in children; endometriosis and uterine fibroids in women; and advanced prostate cancer and sexual deviance in men. The drugs have never been licensed as a treatment for gender dysphoria.

The paper begins by describing how puberty suppression was invented, primarily by the psychologist Peggy Cohen-Kettenis, in the 1990s. It reveals the gap between the protocol described in formal manifestos and actual clinical practice. The second section examines the rationale for this intervention, focusing on two claims—that GnRHa is reversible and that it serves as diagnosis—and two omissions—the association between gender dysphoria and homosexuality and the effect of GnRHa on sexual development. The third section traces the international adoption of the Dutch protocol. The fourth section scrutinizes evidence from an early cohort of 70 adolescents subjected to puberty suppression at the Amsterdam clinic (de Vries et al., 2011, 2014). This cohort provides the only significant evidence that GnRHa followed by cross-sex hormones and surgery results in improved psychological function and reduced gender dysphoria. The evidence is less persuasive than it appears: the number of observations was considerably fewer than 70, the reported reduction in gender dysphoria depended on incommensurable scales, and the outcomes omit one patient who died because puberty suppression dictated a riskier vaginoplasty. The fifth section pursues the British study designed to replicate the Dutch one; it was withheld from publication for some years, presumably because puberty suppression in this sample failed to improve gender dysphoria or psychological functioning. The poor quality of American studies is also noted. The final section evaluates evidence for the side effects of GnRHa. The negative effect on the accrual of bone mass is well studied, while there is increasing evidence for negative effects on cognitive and emotional development and on sexual functioning.

Origins of the Dutch protocol

Transsexualism as a concept emerged in the mid-twentieth century, following the discovery of cross-sex hormones and advances in plastic surgery (Hausman, 1995). Novel physical interventions were justified by the new theoretical construct of "gender identity" invented by American psychologists and psychiatrists, most notably John Money (1994). Gender identity was conceived as developing in infancy (e.g. Green, 1968), but physical interventions for transsexuals under the age of 18 were vanishingly rare. Money in 1973 advised a doctor to prescribe testosterone to a 15-year-old girl and even to consider mastectomy—but he was unusually reckless and there is no evidence that his advice was followed (Gill-Peterson, 2018, pp. 163–164). Specialist clinics for children and adolescents with gender identity problems were founded in Toronto in 1975, in Utrecht in 1987, and in London in 1989. They provided counseling. Cross-sex hormones had to wait until the patient was referred to an adult clinic, at an age ranging from 16 to 18 (Bradley & Zucker, 1990). Surgeries were not performed under the age of 18 (Petersen & Dickey, 1995). Referrals of children were rare. The London clinic—the only specialized clinic for children with gender dysphoria in the United Kingdom—over its first decade accepted an annual average of 14 patients (Di Ceglie, 2018). In its first seven years the Utrecht clinic averaged 9 per year (Cohen-Kettenis, 1994).

Lowering the age of intervention was driven by the founder of the Utrecht children's clinic, Peggy Cohen-Kettenis. She had established herself in the field of gender medicine in the 1980s, presenting research to international conferences of the Harry Benjamin International Gender Dysphoria Association (HBIGDA), which had been formed by clinicians and academics. She eventually became professor of psychology in the Department of Child and Adolescent Psychiatry at University Medical Center Utrecht (Everaerd et al., 2014). She was closely connected to clinicians at VU Medical Center Amsterdam (affiliated with Vrije Universiteit Amsterdam), which housed the country's clinic for adult transsexuals.

Cohen-Kettenis believed that transsexuals would experience better outcomes if they started treatment before adulthood. By the mid-1990s, she was referring some patients aged 16 and 17 to the Amsterdam clinic for endocrinological intervention prior to cross-sex hormones (Cohen-Kettenis, 1994). Males were given an antiandrogen, cyproterone acetate, which prevented erections and caused breast tissue to grow; females were given progestin to stop menstruation (Gooren & Delemarre-van de Waal, 1996). Johanna, for example, "fulfilled all necessary requirements for early treatment": she did not favor girly things (though neither did her sisters), she was fond of soccer, she never dated in school (perhaps not surprising given that she was homosexual), and her parents discovered her wearing a tight t-shirt to conceal her breasts (Cohen-Kettenis et al., 1998, p. 124). Brought to the clinic at 17, she was prescribed progestin for four months and then testosterone. Within two years Jaap (as Johanna had become) underwent mastectomy, hysterectomy, and oophorectomy, and obtained a new birth certificate. Evidence to support such early treatment came from the first 22 patients from the Utrecht clinic, interviewed in their twenties, from one to five years after surgery (Cohen-Kettenis & van Goozen, 1997; Kuiper & Cohen-Kettenis, 1988). They were compared to a larger group of transsexuals who had transitioned later in adulthood in previous decades (Kuiper and Cohen-Kettenis 1988). Her former patients showed better psychological functioning and "more easily pass in the desired gender role" (Cohen-Kettenis & van Goozen, 1997, p. 270). One problem with the comparison is that they had transitioned in a more tolerant era. Another is the fact that they were still young; most had no sexual partner. Moreover they had not reached an age at which they might regret their inability to conceive children. (This group has not since been followed up.) Cohen-Kettenis' initiative was praised by Money: he singled out her contribution to a conference in London as "the bravest" (1998, p. xviii).

Cohen-Kettenis had two collaborators at Amsterdam. One was Henriette Delemarre-van de Waal, a pediatric endocrinologist. She had expertise using the new GnRHa drugs—developed in the 1980s—to treat precocious puberty and other conditions (e.g. Schroor et al. 1995). The other was Louis Gooren, a psychiatrist and endocrinologist who was installed as the world's first professor of transsexuality in 1989. His inaugural professorial lecture was addressed by Cohen-Kettenis and by Money, who flew over from Johns Hopkins University (Nederlands Tijdschrift voor Geneeskunde 1989). Like the pioneering generation who created transsexualism, Gooren saw gender dysphoria as an intersex condition: "there is a contradiction between the genetic, gonadal and genital sex on the one hand, and the brain sex on the other" and therefore "we must provide them with reassignment treatment which meets their needs" (Gooren, 1993, p. 238). This hypothesis was apparently vindicated when he coauthored an article in *Nature* showing that the volume of the central subdivision of the bed nucleus of the stria terminalis in six male-to-female transsexuals was closer to the volume found in females than in males (Zhou et al., 1995). "Unfortunately," as he recently acknowledged, "the research has never been replicated" (Gooren, 2021, p. 50; see also Kreukels & Burke, 2020).

GnRHa was introduced as a treatment for gender dysphoria in two articles. Gooren and Delemarre-van de Waal (1996) proposed the "Feasibility of Endocrine Interventions in Juvenile Transsexuals." More influential was a case study of the first "adolescent transsexual" treated with GnRHa (Cohen-Kettenis and van Goozen 1998). From the age of 5, FG "had made it very clear that I was supposed to be a boy" (FG, 2021, p. 131). It later transpired that FG was sexually attracted to women. FG's father, an Italian with traditional views on gender, disapproved of his daughter's masculinity, and serious conflict ensued. Extensive psychotherapy did not improve matters; FG wrote a suicide note at the age of 12. When FG was 13, Delemarre-van de Waal prescribed triptorelin.[2] Three years later, around 1990, FG came to the Utrecht gender clinic, and Cohen-Kettenis was impressed by FG's "boyish appearance" (Cohen-Kettenis, 2021, p. 115). The clinic provided therapy and introduced FG to other adolescent girls who identified as transsexual. (Whether FG was introduced to any adolescents who identified as lesbian is not recorded.) FG's puberty suppression continued until the age of 18, when testosterone commenced, followed by multiple surgeries: mastectomy, hysterectomy, oophorectomy, and

metaidoioplasty. Awaiting the last surgery at the age of 20, FG was "happy with his life" and "never felt any regrets"; gender dysphoria was apparently cured (Cohen-Kettenis & van Goozen, 1998, p. 247).

Puberty suppression remained exceptional for some years. By 2000, GnRHa had been administered to only 7 children under the age of 16 (Cohen-Kettenis et al., 2000). The new treatment regime was codified at VU Medical Center in Amsterdam, where Cohen-Kettenis was appointed professor of medical psychology in 2002, moving with her clinic. The "Dutch protocol" was published in an influential article in 2006, supported financially by Ferring Pharmaceuticals, the manufacturer of triptorelin (Delemarre-van de Waal & Cohen-Kettenis, 2006, p. S137). GnRHa could be administered to transsexuals as young as Tanner stage 2—marked by the first growth of pubic hair and for girls by budding breasts and for boys by growing testicles—as long as they had reached the age of 12. The adolescent would usually then begin "to live permanently in the role of their desired sex" (Delemarre-van de Waal & Cohen-Kettenis, 2006, p. S132). After some years of puberty suppression, the youth would start cross-sex hormones at the age of 16 and then surgeries at the age of 18. Eligibility criteria for puberty suppression appeared strict. First, gender dysphoria should have begun early in childhood, and dysphoria should have worsened with the onset of puberty. Second, the patient should be psychologically stable, and not suffer from other mental health problems. Third, the patient should have support from their family. As the protocol was formalized, the number of children undergoing puberty suppression increased markedly. Between 2000 and 2008, GnRHa was prescribed to 111 children, about one per month (de Vries et al., 2011). One of them was Valentijn de Hingh, the subject of a television documentary (Nietsch, 2007). After a teacher was disconcerted by the boy's passion for dolls, de Hingh at the age of 5 was diagnosed with gender dysphoria by Cohen-Kettenis (de Hingh, 2021). GnRHa was administered from the age of 12 in 2002.

The protocol as published was not always strictly followed by the clinicians. The minimum age of 12 for puberty suppression was not observed in every case (de Vries, 2010, p. 104). De Hingh had regular endocrinological checkups from the age of 10, presumably so that puberty suppression could commence as soon as Tanner stage 2 was reached. Likewise, cross-sex hormones sometimes started before the age of 16, as young as 13.9 years (de Vries et al., 2011, p. 2278). Family support was not essential, as the clinic administered GnRHa to a 14-year-old—who was institutionalized due to a physical handicap—against the parents' objections (Cohen-Kettenis and Pfäfflin 2003). A British television documentary from the mid-1990s provides a glimpse of actual practice (Morse, 1996). *The Wrong Body* took three English young people to Amsterdam and Utrecht, to see transgender medicine at its most advanced. Fredd Foley, aged 13, met Gooren to learn about puberty suppression; this was around the time it was proposed in the medical literature (Gooren and Delemarre-van de Waal 1996). After returning to England and being refused GnRHa by the London clinic, Foley's mother telephoned Gooren who agreed to write a three-month prescription of triptorelin. "If your child knows for sure he is transsexual," he said, "I would not let puberty happen." Gooren's willingness to prescribe drugs for a child in another country, met briefly in front of the cameras, against the wishes of the child's own psychiatrist, hints that the assessment process was not always as rigorous as portrayed in the published literature. As Cohen-Kettenis said in the documentary, "it's very difficult to give exact criteria, in some cases you have the feeling that the adolescent has thought about it and knows pretty well what she or he is doing."

The Dutch protocol scrutinized

The Dutch protocol comprised not just a drug (GnRHa) and a treatment regime (from age 12 or Tanner stage 2) but also two discursive claims. The first was reversibility. The initial article declared GnRHa to be "fully reversible; in other words, no lasting undesired effects are to be expected" (Gooren & Delemarre-van de Waal, 1996, p. 72). The phrasing hinted at the lack of actual evidence. Suppressing puberty for a short time, on the order of months, might be expected

to have a negligible effect on a child's development. Yet the Dutch protocol entailed suppression for up to four years (from age 12 to 16); for FG it lasted at least five years (from 13 to 18). It was implausible to claim that suppressing puberty for so many years would have no lasting effect if the child were to stop GnRHa and restart their natal sex hormones. On occasion this was acknowledged, as when Delemarre-van de Waal and Cohen-Kettenis' (2006, p. S137) manifesto stated that "It is not clear yet how pubertal suppression will influence brain development." Ten years later, however, Cohen-Kettenis still claimed that puberty suppression was "completely reversible" (Cohen-Kettenis, 2016; see also de Vries et al., 2016). The postulate of reversibility, however implausible, helped to avoid the question of whether a child aged 12 (or below) could give consent to this endocrinological experiment. HBIGDA's Standards of Care warned that cross-sex hormones "are not, or are not readily, reversible" (HBIGDA, 1985, p. 83). By pronouncing GnRHa to be reversible, the Dutch protocol demarcated a boundary between one endocrinological intervention and another.

The second claim was that puberty suppression was a diagnostic tool. The case study of FG described GnRHa as an "aid in diagnosis and treatment" (Cohen-Kettenis & van Goozen, 1998). This echoed the conception of cross-sex hormones as "both therapeutic and diagnostic in that the patient requesting such therapy either reports satisfaction or dissatisfaction regarding the results" (HBIGDA, 1985, p. 85). GnRHa was posited to provide space for therapeutic exploration of gender identity, without the pressure of the physical changes accompanying puberty (Delemarre-van de Waal & Cohen-Kettenis, 2006). This claim was plausible, though it was also plausible that stopping normal cognitive, emotional, and sexual development would impede such exploration. In the event, the Dutch clinicians found that the diagnostic test invariably yielded the same result: "none of the [54] patients who were selected for pubertal suppression has decided to stop taking GnRHa" (Delemarre-van de Waal & Cohen-Kettenis, 2006, p. S136). This might be explained by a rigorous selection process. An alternative explanation is that puberty suppression becomes a self-fulfilling prophecy. Subsequent experience in the Netherlands and in other countries confirms the fact that 96%–98% of children who undergo puberty suppression continue to cross-sex hormones (Brik et al., 2020; Carmichael et al., 2021; Wiepjes et al., 2018).

The framing of GnRHa as diagnostic circumvented a problem recognized in the earliest articles. "Not all children with GID [Gender Identity Disorder] will turn out to be transsexuals after puberty," acknowledged Cohen-Kettenis and Gooren (1999, p. 319). "Prospective studies of GID boys show that this phenomenon is more closely related to later homosexuality than to later transsexualism." They cited three longitudinal studies of feminine boys (Green, 1987; Money & Russo, 1979; Zuger, 1984).[3] The best known is Richard Green's attempt at "studying pretranssexuals" by selecting a group of "sissy boys" (Green, 1987, p. 12). After fifteen years, to his surprise, only one out of 44 was contemplating transsexuality, whereas two thirds had become bisexual or homosexual men. Given such studies, Cohen-Kettenis concluded that "most GID children under 12 will not grow up to become transsexuals" (Cohen-Kettenis & van Goozen, 1997, p. 246). These findings were downplayed in subsequent publications; the key manifestos for the Dutch protocol did not mention homosexuality and did not cite any study of feminine boys (Cohen-Kettenis et al., 2008; Delemarre-van de Waal & Cohen-Kettenis, 2006). The assertion that "GID persisting into early puberty appears to be highly persistent" rested on slender evidence (Cohen-Kettenis et al., 2008, p. 1895). The only relevant cited source described adolescents who had been first assessed at ages ranging from 13 to 18, a range extending well beyond early puberty (Smith et al., 2001). This source did not support the hypothesis that the probability of gender dysphoria persisting to adulthood jumped suddenly on the cusp of age 12, from under 50% to virtually 100%. What is known is that most adolescents subjected to puberty suppression were homosexual. Of the first 70 adolescents referred to the Amsterdam clinic from 2000 to 2008 and given GnRHa, 62 were homosexual while only 1 was heterosexual (de Vries et al., 2011).

The crucial advantage of puberty suppression was creating "individuals who more easily pass in to the opposite gender role" (Delemarre-van de Waal & Cohen-Kettenis, 2006, p. 155). The emphasis was on external appearance, especially height.[4] That word appears 23 times in

Delemarre-van de Waal's review of puberty suppression (Delemarre-van de Waal, 2014). There is one cursory reference to "loss of fertility." The words orgasm, libido, and sexuality do not appear. This is curious because it was well known that men taking GnRHa for prostate cancer experience complete loss of erotic interest (Marumo et al., 1999). The drug is therefore licensed to chemically castrate men with sexual obsessions. Gooren was an early advocate for this usage. He warned that the side effects "may be very uncomfortable" for men with paraphilias (Gijs & Gooren, 1996, p. 279); no such warning accompanied his recommendation of the same drug for adolescents experiencing gender dysphoria. The Dutch clinicians did not ask whether blocking the normal development of erotic desire would affect their patients' understanding of their own body and their interest in future sexual and romantic relationships.

One significant disadvantage of puberty suppression for males was not mentioned in the 2006 manifesto for the Dutch protocol, though it had been raised at a conference in the previous year (GIRES, 2005). Stopping sexual development meant the penis did not grow, and so "the genital tissue available for vaginoplasty may be less than optimal" (Cohen-Kettenis et al., 2008, p. 1895). This made it more likely that the orifice would need to be lined with a portion of the patient's intestine rather than the inverted penis (van de Grift et al., 2020). Out of 49 patients at Amsterdam who started GnRHa at Tanner stage 2 or 3, 71% required intestinal vaginoplasty (van der Sluis et al., 2021). This procedure is more invasive, requiring a second surgical site, and it entails greater risk of complications such as rectal fistula. Surgical techniques have been refined so that the "possible occurrence of intestinal discharge could be kept under control" (Bouman, 2021, p. 141), but one quarter of the patients need further corrective surgeries (Bouman et al., 2016).

International adoption of the Dutch protocol

The Dutch protocol immediately attracted interest in other countries. Cohen-Kettenis and Gooren were already prominent in the field of transgender medicine, exemplified by their election to the Board of Directors of HBIGDA (the former served two four-year terms from 1995 and 2003, while the latter served one term from 1999). Puberty suppression soon entered HBIGDA's Standards of Care in the Sixth Version, approved in 2001. It closely followed the Dutch protocol, but did not specify any minimum age. It was "recommended that the adolescent experience the onset of puberty in his or her biologic sex, at least to Tanner stage Two," while also allowing earlier intervention on the recommendation of more than one psychiatrist (HBIGDA, 2001, p. 10). Recall that the published evidence for the benefits of puberty suppression then comprised a single case study of one patient—FG—awaiting final surgery.

In the United States, adoption was led by Norman Spack, a pediatric endocrinologist. More than once he recalled "salivating" at the prospect of treating patients with GnRHa (Hartocollis 2015; Spack 2008, xi). In 2007 he cofounded the Gender Management Service at Boston Children's Hospital, which was the first dedicated clinic for transgender children in America. Its program was based on the Dutch model; the hospital sent a psychologist to Amsterdam to be trained by Cohen-Kettenis (Tishelman et al., 2015). From the outset the Boston clinic offered GnRHa at Tanner stage 2 or 3 with no minimum age (Spack et al. 2012). Spack joined Cohen-Kettenis, Gooren, and Delemarre-van de Waal on the Endocrine Society's committee tasked with writing their first clinical guidelines for "transsexual persons," which recommended GnRHa for children at Tanner stage 2 or 3 (Hembree et al., 2009). "There was an attitudinal shift to be able to say that the Endocrine Society supports this," he later recalled (Ruttimann, 2013, p. 19). The shift is visible in data from 43 children's hospitals on prescriptions of one GnRHa drug (histrelin acetate): it was never prescribed for gender dysphoria between 2004 and 2009 and was then prescribed to 92 patients from 2010 to 2016, most in the final years of the period (Lopez et al., 2018).

Oprah Winfrey Television broadcast the documentary *I Am Jazz: A Family in Transition* in 2011 (Stocks, 2011). Its dramatic structure was similar to *The Wrong Body*, focusing on the

looming threat of puberty as Jazz Jennings reached the age of 11. Jennings had been diagnosed with gender dysphoria at the age of 3 and had appeared on national television at the age of 7, when the family created the TransKids Purple Rainbow Foundation (Jennings & Jennings, 2016). The documentary showed the family consulting with a pediatric endocrinologist, who confirmed that Tanner stage 2 had been reached. The denouement was not shown, but Jenning's mother was clear: "you have to kinda nip puberty in the bud, you want to block it" (Stocks, 2011). Jennings did indeed commence puberty suppression some months later. The number of clinics for "gender-nonconforming children and adolescents" multiplied, and within a few years 32 of them advertised puberty blockers (Hsieh & Leininger, 2014).

England provides an example of adoption driven by patients rather than clinicians. *The Wrong Body* had promoted the Dutch approach to 3 million viewers (Nataf, 1999). Dissatisfaction at the cautious policy of the London clinic—still headed by its founder, Domenico Di Ceglie—became increasingly vocal. Sustained pressure came from the parents of children who identified as transgender, organized in the Gender Identity Research and Education Society (GIRES) and Mermaids. GIRES obtained funding from medical charities to organize an international symposium in London in 2005 to develop consensus guidelines for endocrinological intervention, which was attended by Cohen-Kettenis, Delemarre-van de Waal, and Spack. GIRES (2006) warned that "those who can in any way afford to do so have to consider taking their children to the USA." The first was Susie Green, later the chief executive of Mermaids. In 2007 she took her son Jackie, aged 12, to Boston to obtain GnRHa from Spack (Sloan, 2011). A presentation at Mermaids instructed parents in this medical tourism (Mermaids, 2007). Spack treated seven more British children over the next few years (Glass, 2012). The conflict between parents and clinicians climaxed in 2008, with two clashing conferences. The Royal Society of Medicine organized a meeting on adolescent gender dysphoria, which drew criticism for the lack of overseas speakers advocating for puberty blockers, even though it had invited Delemarre-van de Waal. The cofounder of GIRES, whose child had transitioned in their late teens two decades earlier, used the new epithet "transphobic" to describe the cautious clinicians (Groskop, 2008). Richard Green—the author of *Sissy Boys*, then in London as a visiting professor—quickly organized a rival conference to demand puberty suppression (Green, 2008). Speakers included the usual cast of clinicians, including Spack, and also patients and their parents, including two Dutch transgender adolescents. The demand for puberty suppression was becoming irresistible.

Di Ceglie was soon replaced as director of the London clinic (renamed the Gender Identity Development Service and located at the Tavistock and Portman NHS Foundation Trust) by Polly Carmichael, a clinical psychologist. The clinic in 2011 began to offer GnRHa from the age of 12, initially as part of an experimental study (Biggs, 2019b, 2019c). Before any outcomes were published, Carmichael declared success: "Now we've done the study and the results thus far have been positive we've decided to continue with it" (Manning and Adams, 2014). She even appeared on BBC Children's Television to promote puberty suppression, in a documentary about a 13-year-old girl who wanted to be a boy, Leo. Carmichael reassured Leo about GnRHa: "the good thing about it is, if you stop the injections, it's like pressing a start button and the body just carries on developing as it would if you hadn't taken the injection" (Niland, 2014). In 2015 the National Health Service adopted a policy of offering GnRHa for adolescents at Tanner stage 2, without age restriction (NHS England, 2015).

Evidence from the Amsterdam clinic

By the mid-2010s, then, the Dutch protocol was established as the standard for transgender medicine. It was apparently vindicated when longitudinal data was published on a cohort of 70 adolescents referred to the clinic between 2000 and 2008 and then subjected to puberty suppression. The lead author, Annelou de Vries, received her doctorate under the supervision of Cohen-Kettenis. Outcomes were initially measured as the patient was transitioning from GnRHa to cross-sex hormones, at ages ranging from 14 to 19. "Behavioral and emotional problems and

depressive symptoms decreased, while general functioning improved" (de Vries et al., 2011, p. 2276). Outcomes were subsequently measured soon after the patient's final surgery (vaginoplasty or mastectomy and hysterectomy with oophorectomy), at ages ranging from 19 to 22. The authors concluded that "gender dysphoria had resolved, psychological functioning had steadily improved, and well-being was comparable to same-age peers" (de Vries et al., 2014, p. 696).

When scrutinized, however, the evidence is less persuasive. The sample was small: final outcome measures were available for subsets of patients numbering between 32 and 55. The finding that gender dysphoria had resolved depended on the Utrecht Gender Dysphoria Scale and the Body Image Scale, which have separate questionnaires for each sex. The researchers switched versions over the course of the study (Levine et al., 2022). A boy who wanted to become a girl, for example, answered the male questionnaires at baseline before puberty suppression, and then the female versions following surgery—so would be rating agreement with the statement "I hate menstruating because it makes me feel like a girl" (C. Schneider et al., 2016) and satisfaction with "ovaries-uterus" (Lindgren & Pauly, 1975). The inclusion of such meaningless questions compromises the measurement of change in gender dysphoria. The results after surgery exclude eight patients who refused to participate in the follow-up or were ineligible for surgery, and one patient killed by necrotizing fasciitis during vaginoplasty. The authors did not mention the fact that this death was a consequence of puberty suppression: the patient's penis, prevented from developing normally, was too small for the regular vaginoplasty and so surgery was attempted with a portion of the intestine, which became infected (Negenborn et al., 2017). A fatality rate exceeding 1% would surely halt any other experimental treatment on healthy teenagers.

One inevitable limitation of the study was the measurement of results soon after surgery, which repeated the problem with the first study of adolescent transsexuals (Cohen-Kettenis & van Goozen, 1997). As Cohen-Kettenis notes, "a truly proper follow-up needs to span a minimum period of 20 years" (Cohen-Kettenis, 2021, pp. 117–118). A subsequent follow-up of this cohort is in preparation (Bazelon, 2022). The only long-term outcome published in the literature is that of the very first patient, FG, who was followed up again at the age of 35. FG did not regret transition, but scored high on the measure for depression. Owing to "shame about his genital appearance and his feelings of inadequacy in sexual matters," he could not sustain a romantic relationship with a girlfriend (Cohen-Kettenis et al., 2011, p. 845). Ironically, a "strong dislike of one's sexual anatomy" is one of the diagnostic criteria for gender dysphoria in children (according to *DSM*-5), and so in this respect it is not clear how the dysphoria had been resolved. The clinicians were more interested in FG's height, which they compared punctiliously to the Italian as well as the Dutch height distribution. Cohen-Kettenis concluded that "the negative side effects are limited" (Cohen-Kettenis et al., 2011, p. 843). Delemarre-van de Waal's (2014, p. 194) summary was even more optimistic: "He was functioning well psychologically, intellectually, and socially." Now aged 48, FG has given two recent interviews. FG's situation seems to have improved, and he now has a serious girlfriend. He describes puberty suppression as "life-saving" in his case (FG, 2021, p. 132) but also recommends that it should require a significant assessment process (Bazelon, 2022). In a recent interview, Valentijn de Hingh, who at the age of 31 now identifies as non-binary, emphasizes that "diagnosis and treatment at a young age were not wrong." At the same time, de Hingh wonders "wasn't that very young? To have been seeing a psychologist, having been examined and diagnosed from the age of five" (de Hingh, 2021, p. 182).

Replicating the Dutch results

An international study of puberty suppression—involving London and Boston as well as Amsterdam—was first mooted in 2005 (GIRES, 2005). The Boston clinic dropped out, but eventually an experiment along Dutch lines was begun in London in 2010. The entry criteria were "consistent with the protocol used at the Amsterdam Gender Clinic" (Viner et al., 2010, p. 6) and the outcome measures replicated those used by the Amsterdam longitudinal study (de

Vries et al., 2011, 2014). From 2011 to 2014, 44 adolescents aged from 12 to 15 years commenced puberty suppression. Outcomes for all subjects after two years on GnRHa were thus collected by 2016. Preliminary results were presented to the World Professional Association for Transgender Health (as HBIGDA had been renamed) in Amsterdam. In her keynote address, Carmichael observed that "our results have been different to the Dutch" (Carmichael, 2016). According to one presentation, adolescents after one year of GnRHa "report an increase in internalising problems and body dissatisfaction, especially natal girls" (Carmichael et al., 2016). Another presentation was also negative: "Expectations of improvement in functioning and relief of the dysphoria are not as extensive as anticipated, and psychometric indices do not always improve nor does the prevalence of measures of disturbance such as deliberate self harm improve" (Butler, 2016). These conference papers were not published as articles, following the typical fate of medical experiments that fail to produce positive results (Johnson & Dickersin, 2007).

Instead, the London clinic published an article claiming that "adolescents receiving also puberty suppression had significantly better psychosocial functioning after 12 months of GnRHa … compared with when they had received only psychological support" (Costa et al., 2015, p. 2206). The group subjected to puberty suppression were aged between 13 and 17, and must have included some of the 44 experimental subjects. This group comprised 101 adolescents at the outset, diminishing to 35 after twelve months. This high rate of attrition was not explained in the article. Anyway, the data showed no statistically significant difference between the group given GnRHa and counseling and the group given only counseling (Biggs, 2019a).

The full outcomes from the experiment were published following a protracted campaign involving publicity in newspapers and television (e.g. Tominey & Walsh, 2019), complaints to the ethics committee which approved the research (Health Research Authority, 2019), a Parliamentary question (Blackwood of North Oxford, 2019), and a judicial review (Keira Bell and Mrs A v Tavistock NHS Trust, 2020). Out of the 44 subjects in the experiment, all but one transitioned to cross-sex hormones. Outcomes were taken after 12 months of puberty suppression for all patients, and after 24 months for the subset waiting to reach the age of 16 when they could start cross-sex hormones. The headline finding was that "GnRHa treatment brought no measurable benefit nor harm to psychological function in these young people," and gender dysphoria likewise did not improve (Carmichael et al., 2021, p. 20). This is all the more surprising because a placebo response would be expected in patients who had volunteered to pioneer this intervention in Britain (Kirsch, 2019). There was no disaggregation by sex, which is unfortunate because outcomes were evidently worse for natal girls than for boys (Biggs, 2020; Carmichael et al., 2016).

The researchers did not compare their findings to the outcomes from the Amsterdam clinic after puberty suppression (de Vries et al., 2011). Comparison is undertaken here, using available data on two question batteries.[5] The Youth Self-Report (YSR) enables the adolescent to describe their problems, while the Child Behavior Checklist (CBCL) provides a parent's assessment. YSR and CBCL each yield three *T*-scores: one for Internalizing Problems like anxiety; one for Externalizing Problems like anger; and a Total Problem score, combining these two along with other problems such as social isolation (Achenbach & Rescorla, 2001). *T*-scores are normalized relative to reference scores (for males and for females aged 12–18), with a mean of 50 and standard deviation of 10. The Amsterdam clinic reported these measures for 54 subjects, compared to 41 for the London clinic. The two samples were similar at the outset of puberty suppression: the mean age at Amsterdam was 14.8, the median at London was 13.6; females comprised 53% of the Amsterdam sample, 43% of the London one. Figure 1 depicts the mean scores at baseline before the commencement of puberty suppression, along with the 95% confidence interval. There was no discernible difference between the Amsterdam and London samples in any component of CBCL or YSR. At the Amsterdam clinic, the subjects completed the questionnaires again when they transitioned to cross-sex hormones, after a mean of 1.9 years. At the London clinic, the questionnaires were completed at 12-month intervals, and so I take the latest available before the end of puberty suppression; the mean duration is 1.4 years. Figure 2

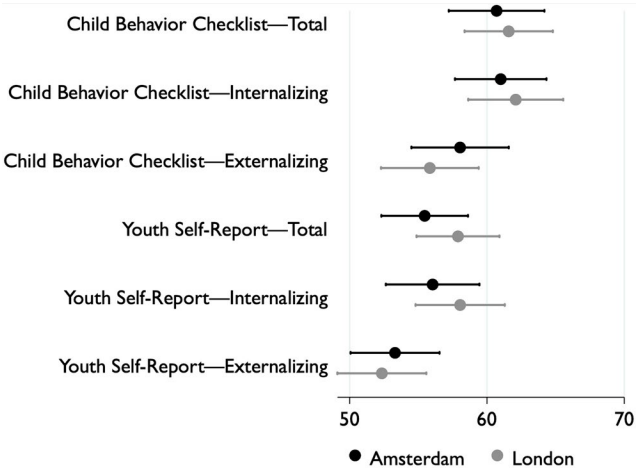


Figure 1. Psychological functioning before puberty suppression with GnRHa. The circle shows the mean *T*-score at baseline. The line traces the 95% confidence interval. $N = 54$ at Amsterdam, 41 at London. Data from de Vries et al. (2011, Table 2) and Carmichael et al. (2021).

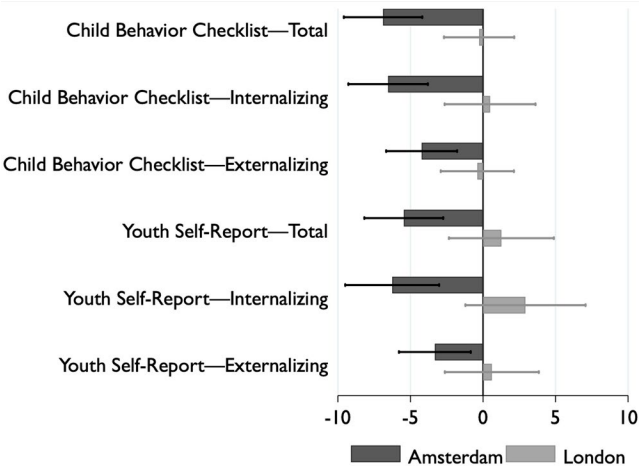


Figure 2. Change in psychological functioning after puberty suppression with GnRHa. The bar shows the change in *T*-score from baseline; negative values indicate reduced problems. The line traces the 95% confidence interval. $N = 54$ at Amsterdam, 41 at London. Data reported from de Vries et al. (2011, Table 2) and Carmichael et al. (2021).

shows how the scores changed since baseline. The Amsterdam sample improved—fewer problems were reported by the subjects and their parents—on all six measures (p = .000004 … .003). The London sample, by contrast, experienced no discernible change (p = .16 … .82). With one exception (YSR Externalizing Problems), the differences between the change in Amsterdam and the change in London are statistically significant (p = .0006 … .03, assuming equal variance).

The London clinic's failure to replicate the positive results found by the Amsterdam clinic after puberty suppression demonstrates that the Dutch results cannot be extrapolated to other countries. The reason for the failure to replicate could perhaps lie in the quality of care offered by the clinics or in the characteristics of their patients. Although the two samples had indistinguishable baseline scores on YSR and CBCL, on another measure of psychological functioning—the Children's Global Assessment Scale (CGAS), which is scored by the clinician—the adolescents attending the London clinic were significantly worse at the outset. This fits the general pattern in adolescents referred to European gender clinics: those at Amsterdam have fewer psychological problems and better peer relationships than those at London (de Graaf et al., 2018). The failure

to replicate could simply exemplify a general phenomenon in medicine (and science generally): a large effect found in a nonrandomized study with a small sample usually either declines in magnitude or disappears altogether in subsequent studies (e.g. Ioannidis, 2005). Given the London clinic's failure to find favorable results after puberty suppression, it has no incentive to follow up the 43 subjects who transitioned to cross-sex hormones and potential surgery. It loses track of all its patients after the age of 18, blaming "the frequent change in nominal and legal identity, including NHS number in those referred on to adult services" (Butler et al., 2018, p. 635).

One other clinic has published a comparable longitudinal study of puberty suppression. The Hamburg Gender Identity Service followed 11 adolescents who were administered GnRHa for an average of one year, but such a tiny sample provides insufficient statistical power for any conclusions (Becker-Hebly et al., 2021). Three American studies of puberty suppression have been published: from Stony Brook (Achille et al., 2020), Dallas (Kuper et al., 2020), and Seattle (Tordoff et al., 2022).[6] None tried to replicate the Amsterdam and London longitudinal studies, choosing completely different measures, with one exception (BIS is used by Kuper et al., 2020). Each introduced a different set of measures: Quick Inventory of Depressive Symptoms, Screen for Child Anxiety Related Emotional Disorders, Center for Epidemiologic Studies Depression Scale, Quality of Life Enjoyment and Satisfaction Questionnaire, Generalized Anxiety Disorder 7-item scale, and the Patient Health Questionnaire 9-item scale. The last scale was common to two studies, but even they were not comparable: one used the version for teenagers, the other the adult version which the researchers chose to dichotomize. All the samples were tiny: 19, 23 (including an unspecified number of males given anti-androgens and females given medroxyprogesterone rather than GnRHa), and 25. Results were reported inconsistently: sometimes the outcomes for the sample subjected to puberty suppression were combined with a much larger sample on cross-sex hormones; sometimes the parameters of complex multivariate models were reported while the within-subject change during puberty suppression was concealed (Singal, 2022). Finally, some results were vitiated by high—and unexplained—rates of attrition: 47% of the subjects in one study were excluded because they failed to fill in the questionnaires at three points in time (Achille et al., 2020). What is frustrating is that if these researchers had simply followed the methods of de Vries et al. (2011), these three small samples would have contributed to cumulative knowledge. Finally, a large-scale American study recruited 90 subjects for puberty suppression—from Boston, Chicago, Los Angeles, and San Francisco—between 2016 and 2018 (Olson-Kennedy et al., 2019). Outcomes after 24 months have evidently been collected, but only baseline results have been published (Chen et al., 2021).

Evidence on side effects

On the side effects of puberty suppression, there is most evidence on bone density. That GnRHa would cause "an insufficient formation of bone mass" was initially dismissed "of no great concern" (Gooren & Delemarre-van de Waal, 1996, p. 72). Then it was recognized that patients could "end with a decreased bone density, which is associated with a high risk of osteoporosis" (Delemarre-van de Waal & Cohen-Kettenis, 2006, p. S134). The detrimental effect of GnRHa on the accrual of normal bone mass has been documented in several longitudinal studies from the Amsterdam clinic (Klink et al., 2015; Schagen et al., 2020; Stoffers et al., 2019; Vlot et al., 2017), the London clinic (Biggs, 2021; Joseph et al., 2019), and a clinic in Ottawa (Navabi et al., 2021). Less obviously, adolescents who seek GnRHa for gender dysphoria have a lower distribution of bone density compared to the population of the same sex and age (see also Lee et al., 2020). This reflects in part the high prevalence of eating disorders.

Bone mineral density (BMD) is measured by a dual energy X-ray absorptiometry scan over the spine and the hip. The absolute value of BMD is standardized as a *Z*-score, expressing this individual's BMD relative to the population of the same sex and age. BMD can be adjusted for height to derive the volumetric bone mineral apparent density (BMAD), which is likewise standardized as a *Z*-score. A *Z*-score below −2 is considered low; it indicates bone density in the

lowest 2.3% of the population. The salience of this threshold is revealed by the London clinic's protocol which required both spine and hip *Z*-scores to exceed −2 to be eligible for GnRHa (Viner et al., 2010). This was subsequently relaxed "in exceptional circumstances" if clinicians "feel that on the balance of risks, pubertal suppression is an appropriate option despite risks of osteoporosis in later adult life" and patients "understand the risks of GnRH analogue treatment for bone density (i.e. risks of later osteoporosis)" (Viner et al., 2012).

Most studies of bone density after puberty suppression summarize the distribution of *Z*-scores by mean and standard deviation; only two provide information on the lower tail of the distribution, which is what matters clinically. At the Amsterdam clinic, 56 transgender adolescents were treated with GnRHa, commencing at ages ranging from 11 to 18, for an average duration of 1.7 years. After puberty suppression, the minimum *Z*-score for spine BMAD was −2.4, and the minimum hip BMAD was −3.4 (Vlot et al., 2017). Normally we would expect to find a *Z*-score below −3 in only 0.13% of the population—1 in 741. At the London clinic, 24 adolescents were treated with GnRHa, commencing at ages ranging from 12 to 14, for a duration of 24 months. After puberty suppression, the hip BMD *Z*-score was below −2 for 7 patients. The spine BMD *Z*-score was below −2 for 7 patients, including 4 patients with *Z*-score below −3; the spine BMAD *Z*-score was below −2 for 8 patients, including 3 with *Z*-score below −3 (Biggs, 2021). Clearly, then, a significant minority of patients have abnormally low bone density after puberty suppression. The subsequent administration of cross-sex hormones increases bone mass, but *Z*-scores remain below the baseline recorded at the outset of puberty suppression (Klink et al., 2015; Stoffers et al., 2019; Vlot et al., 2017), with the possible exception of females who take testosterone after starting GnRHa early in puberty (Schagen et al., 2020).

What is not clear is the consequence of abnormally low bone density. Information on fractures, for example, has been published only for adults taking cross-sex hormones who had not undergone puberty suppression (Wiepjes et al., 2020). Anecdotally, a female patient at the London clinic who started GnRHa at age 12 then experienced four broken bones by the age of 16 (Bannerman, 2019). A Swedish television documentary discovered one female who was given GnRHa from age 11 to 15 by the Karolinska University Hospital in Stockholm, and now suffers from severe osteoporosis, including continual skeletal pain (SVT, 2022). This case—along with two others whose puberty suppression was terminated following concerns about bone density—led Sweden to restrict the use of GnRHa for adolescents with gender dysphoria.

The effects of puberty suppression on emotional and cognitive development are more difficult to ascertain, but more consequential as they could potentially affect the capacity to consent to cross-sex hormones and surgery. One case report of puberty suppression commencing just before age of 12 measured a drop in IQ by 10 points after 28 months (M. A. Schneider et al., 2017). A single case is insubstantial, of course, but there are similar findings from children treated with GnRHa for precocious puberty. A study of 25 children measured a drop of 7 points after two years (Mul et al., 2007); another study found a gap of 8 points between 15 treated children and a matched control group (Hayes, 2017; Wojniusz et al., 2016). Unfortunately the Amsterdam clinic's longitudinal study of puberty suppression measured IQ only at baseline and did not measure it again (de Vries et al., 2011, 2014). A small study from the clinic found that 8 adolescent males undergoing puberty suppression performed worse in a test of executive functioning than three control groups; the differences are statistically significant, but the samples are small (Staphorsius et al., 2015). Randomized control trials of non-human animals provide evidence of the substantial effects of puberty suppression. In sheep, GnRHa impairs spatial memory, and this effect remains after the treatment is stopped—thus demonstrating the irreversibility of puberty suppression (Hough et al. 2017a; 2017b). Counterintuitively, GnRHa also leads to greater differences between males and females in foraging behavior (Wojniusz et al., 2011). In mice, the effects of GnRHa vary by sex: males develop stronger preference for other males and an increased stress response; females exhibit increased anxiety and despair-like behavior (Anacker et al., 2021).

Even less is known about the effects of puberty suppression on sexual functioning. Jennings, who started on GnRHa at the age of 11, has no libido and cannot orgasm. Jennings' surgeon,

Marci Bowers, who has performed over 2,000 vaginoplasties, acknowledges that "every single child … who was truly blocked at Tanner stage 2, has never experienced orgasm. I mean, it's really about zero" (Bowers, 2022). This remark refers to males. The effects of puberty suppression at such an early stage on females is unknown. FG is reportedly able to orgasm (de Vries et al., 2011). One patient at the London clinic who took GnRHa from the age of 12 to 16 but did not continue to cross-sex hormones has experienced no sexual desire in the two years since ceasing GnRHa (Bannerman, 2022). According to de Vries, orgasm is "a very interesting and so far not studied question" (Klotz, 2022).

Conclusion

The use of GnRHa to suppress puberty has proved more popular than could have been envisaged in the mid-1990s. It has become the international standard for treating gender dysphoria and has attracted increasing numbers of patients. Down to 2015, the Amsterdam clinic administered GnRHa to a total of 333 youth aged under the age of 18 (Wiepjes et al., 2018). From 2012 to 2020, the London clinic administered GnRHa to 344 children under the age of 15. Both clinics were overwhelmed by referrals from the mid-2010s, and the lengthening waiting lists provided scope for unscrupulous commercial operations. GenderGP, for example, is a company registered in Singapore and owned by a Welsh doctor which will diagnose a 9-year-old with gender dysphoria over video and prescribe GnRHa on the same day (Medical Practitioners Tribunal Service, 2022). The total number of patients subjected to puberty suppression, worldwide, must run to several thousand. The proponents of GnRHa never reassessed the rationale for the intervention as the numbers multiplied. It is one thing to assert that very rare cases of extreme gender dysphoria—one per year in the Netherlands in the late 1990s—should be treated as juvenile transsexuals. It is another to make this claim for numerous adolescents—currently over a hundred a year in the Netherlands. Given the fact that gender dysphoria lacks an objective diagnosis, the potential for puberty suppression is expansive. When a recent survey in one American school district found 7% of students identifying as "gender diverse," the authors urged that all receive "access to gender affirming care," which in effect means giving GnRHa on request (Kidd et al., 2021, p. 3).[7]

Whether the availability of puberty suppression has increased demand is a question that should be raised. Taking GnRHa early in puberty promises a more passable resemblance to the opposite sex, and this is why it proved so fascinating to television audiences. It is no coincidence that media coverage of transgender youth focuses on those who suppressed puberty at a young age, most famously Jennings. Positive media coverage is known to increase referrals to gender clinics, at least over the short term (Indremo et al., 2022; Pang, de Graaf, et al., 2020). Although Dutch clinicians advise against "a complete social transition … before the very early stages of puberty" (de Vries & Cohen-Kettenis, 2012, pp. 308–309), the availability of GnRHa now makes it feasible for parents to treat a young child as the opposite sex, which guarantees that the child will experience the onset of puberty as catastrophic and thus demand endocrinological intervention. For boys, social transition prior to puberty is a powerful predictor of gender dysphoria persisting into adolescence, even controlling for the degree of dysphoria in childhood (Steensma et al., 2013). This pathway is illustrated by interviews with 30 British parents who had started raising their children as the opposite sex between the ages of 3 and 10. According to one parent, "If you don't give a child puberty-blockers there is a consequence—it's not that nothing happens. There's a massive consequence" (Horton, 2022, p. 13). Another candidly described their child's attitude to counseling at the gender clinic: "at the end of the day, he's just gonna say whatever it is, that makes you shut up, so that he can get the blocker" (Horton, 2022, p. 14).

What has happened to the majority of children with gender dysphoria who used to grow up into gay or lesbian adults? The original articles promoting GnRHa (Cohen-Kettenis & van Goozen, 1998; Gooren & Delemarre-van de Waal, 1996) hypothesized that children whose dysphoria persisted to the age of 12 were destined to become transsexual. This arbitrary age threshold

was soon forgotten. Outside the Netherlands, GnRHa was adopted with no minimum age, and has been prescribed to children as young as 8 years old.[8] Delemarre-van de Waal eventually advocated for GnRHa to be administered before Tanner stage 2, "right at the onset of puberty," followed quickly by cross-sex hormones (Delemarre-van de Waal, 2014, p. 202). And of course the transsexual pathway now begins long before puberty, with social transition and psychological diagnosis. As de Hingh observes, "a diagnosis says you've got a problem that needs to be treated … The medical process, with pills and protocols, takes over the normal process of identification formation" (de Hingh, 2021, pp. 182–183). Clinicians need to explain how they are sure that some of the adolescents being prescribed GnRHa would not have grown into gay or lesbian adults, with their sexuality and fertility intact.

The article that introduced puberty suppression to the medical literature was accurately titled: this endocrinological intervention is designed for juvenile transsexuals (Gooren & Delemarre-van de Waal, 1996). This fact should not be obscured by claiming that puberty suppression is reversible and diagnostic. It is not diagnostic because over 95% of adolescents given GnRHa will continue to cross-sex hormones, and this fraction has not declined even as the number of youths subjected to GnRHa has multiplied by two orders of magnitude. The claim for reversibility was contradicted from the outset by the unknown effect of puberty suppression on brain development. Irreversibility has now been demonstrated by randomized control trials in non-human animals. The central justification for puberty suppression was that it increases outward resemblance to the opposite sex and requires less surgical intervention. Paradoxically, however, early puberty suppression for males will most likely make subsequent genital surgery more risky—this is what killed one of the initial Dutch cohort—with worse results.

Evidence for the benefits of puberty suppression must be acknowledged as slender. Decisions made by clinicians have prevented the collection of robust evidence. The Dutch proponents of GnRHa chose not to conduct a randomized control trial, giving two reasons (de Vries et al., 2011). Firstly, adolescents would have refused to participate, which does not make sense unless they could have obtained GnRHa from another source. Secondly, it would have been unethical to withhold GnRHa from the control group, because the clinicians believed the treatment to be beneficial—this rationale is circular because discovering whether a treatment is truly beneficial requires a randomized control trial. A lesson can be drawn from the use of GnRHa to pause precocious puberty. This was supposed to mitigate short stature, as was apparently shown by small uncontrolled studies (Hayes, 2016), but this effect was called into question by a randomized control trial (Cassio et al., 1999). When the London clinic designed a study to replicate the findings from Amsterdam, the same reasons for avoiding a randomized control study were repeated, along with an argument that subjects would soon realize whether they were receiving treatment or placebo (Viner et al., 2010). Yet this had been no impediment to the trial for children with early puberty.

The decision to rely on uncontrolled studies was exacerbated by other decisions. The Dutch clinicians chose incommensurable scales to measure gender dysphoria, which calls into question their finding that dysphoria declined following cross-sex hormones and surgery. Worse still, American clinicians eschewed the measures of psychological functioning used by the Amsterdam and London clinics (YSR, CBCL, and CGAS), thus ensuring that their tiny samples could not contribute to cumulative knowledge. One final point to remember in evaluating published studies is that the field of transgender medicine is subject to the same publication bias as every other field: unsuccessful results will not be published. This bias is illustrated by the London clinic's attempt to replicate the Amsterdam clinic's findings: the lack of improvement on GnRHa appeared in print only after the clinic was taken to the High Court of Justice for England and Wales.

While the use of GnRHa to suppress puberty helped to create the juvenile transsexual, it could now be creating another "new way of being a person" (Wren, 2020): a sexless adult. This follows from the premise that natal puberty can be a kind of disease, and therefore failure to prevent an "irreversible development of secondary sex characteristics … may be considered unethical" (de Vries et al., 2011, p. 2282). Although the Dutch protocol envisages GnRHa as a

preparatory phase before cross-sex hormones—imagined as undergoing puberty of the opposite sex—the logical conclusion is that hormones of either sex can be treated as vectors of disease. An Australian girl, Phoenix, was socially transitioned into a nonbinary identity at the age of 5 and took GnRHa from age 11. Reaching the age of 16, Phoenix refused to take testosterone because "remaining in an androgynous, peripubertal state is the only way their body can truly reflect their non-binary gender identity" (Notini et al., 2020, p. 743). The clinicians agreed to provide perpetual puberty suppression, despite the known deleterious physical effects—most obviously on bone density—and despite the unknown effects on emotional and cognitive development—which would affect Phoenix's capacity to consent. Phoenix is not the only individual seeking indefinite puberty suppression (Pang, Notini, et al., 2020). Such cases are still exceptional. But cases like FG also used to be exceptional.

Notes

1. The literature sometimes refers to GnRH (or LHRH) analogues, which is a broader classification comprising antagonists as well as agonists.
2. The pediatric endocrinologist was not named in the original article, but her identity is clear from later sources (e.g. Delemarre-van de Waal, 2014). FG is known as "B" in the published literature.
3. Bailey and Zucker (1995) had by then reviewed four additional prospective studies in the same vein as well as numerous retrospective ones. Later prospective studies demonstrated that girls who manifested cross-gender behavior as infants were also more likely to grow up as lesbian, though the association was weaker than for boys (e.g. Li et al., 2017).
4. Pediatric endocrinology's obsession with height has motivated the use of artificial estrogen to accelerate puberty in girls judged as too tall (Cohen & Cosgrove, 2009) and the use of GnRHa to delay puberty in girls judged as too short (Hayes, 2016).
5. A previous comparison (Biggs, 2020) included only 30 subjects from the London clinic and measured outcomes after 12 months. The Stata do-file is posted on Harvard dataverse at https://doi.org/10.7910/DVN/QPRCR1.
6. De Vries (2022) also cites a study from Kansas City (Allen et al., 2019) which includes an unknown number of children subjected to GnRHa before cross-sex hormones, but it took no baseline measure before puberty suppression.
7. The authors calculate the "gender diverse" proportion as 9% because they omit students who skipped the question (Kidd et al., 2021). It is more plausible to include the latter in the denominator, which yields 7%.
8. The London clinic referred a 7-year-old for endocrinological intervention, but it is not known whether GnRHa was actually prescribed before she turned 8 (Butler et al., 2022).

Acknowledgements

Special thanks are due to Sherena Kranat, Susan Matthews, Ema Abbruzzese, and a Dutch detransitioner.

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Achenbach, T. M., & Rescorla, L. (2001). *Manual for the ASEBA school-age forms and profiles: An integrated system of multi-informant assessment*. Burlington, VT: ASEBA.

Achille, C., Taggart, T., Eaton, N. R., Osipoff, J., Tafuri, K., Lane, A., & Wilson, T. A. (2020). Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: Preliminary results. *International Journal of Pediatric Endocrinology*, *2020*(8). doi:10.1186/s13633-020-00078-2

Allen, L. R., Watson, L. B., Egan, A. M., & Moser, C. N. (2019). Well-being and suicidality among transgender youth after gender-affirming hormones. *Clinical Practice in Pediatric Psychology*, *7*, 302–311. doi:10.1037/cpp0000288

Anacker, C., Sydnor, E., Chen, B. K., LaGamma, C. C., McGowan, J. C., Mastrodonato, A., … Denny, C. A. (2021). Behavioral and neurobiological effects of GnRH agonist treatment in mice: Potential implications for puberty suppression in transgender individuals. *Neuropsychopharmacology*, *46*, 882–890. doi:10.1038/s41386-020-00826-1

Bailey, J. M., & Zucker, K. J. (1995). Childhood sex-typed behavior and sexual orientation: A conceptual analysis and quantitative review. *Developmental Psychology*, *31*, 43–55. doi:10.1037/0012-1649.31.1.43

Bannerman, L. (2019, July 26). Puberty blocking drugs: 'For the past four years I've been stuck as a child'. *Times*. https://www.thetimes.co.uk/article/transgender-children-puberty-blocking-drugs-for-the-past-four-years-i-ve-been-stuck-as-a-child-5s6tkh7z2

Bannerman, L. (2022, June 17). 'My puberty was chemically delayed: I was their guinea pig'. *Times*. https://www.thetimes.co.uk/article/my-adolescence-was-chemically-delayed-i-was-their-guinea-pig-bbs3w00ph

Bazelon, E. (2022, June 15). The battle over gender therapy. *New York Times*. https://www.nytimes.com/2022/06/15/magazine/gender-therapy.html

Becker-Hebly, I., Fahrenkrug, S., Campion, F., Richter-Appelt, H., Schulte-Markwort, M., & Barkmann, C. (2021). Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: A descriptive study from the Hamburg Gender Identity Service. *European Child & Adolescent Psychiatry*, *30*, 1755–1767. doi:10.1007/s00787-020-01640-2

Biggs, M. (2019a). A letter to the editor regarding the original article by Costa et al: Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. *Journal of Sexual Medicine*, *16*, 2043. doi:10.1016/j.jsxm.2019.09.002

Biggs, M. (2019b). Britain's experiment with puberty blockers. In M. Moore & H. Brunskell-Evans (Eds.), *Inventing Transgender Children and Young People* (pp. 40–55). UK: Cambridge Scholars Publishing.

Biggs, M. (2019c). *The Tavistock's experiment with puberty blockers*. http://users.ox.ac.uk/~sfos0060/Biggs_ExperimentPubertyBlockers.pdf

Biggs, M. (2020). Gender dysphoria and psychological functioning in adolescents treated with GnRHa: Comparing Dutch and English prospective studies. *Archives of Sexual Behavior*, *49*, 2231–2236. doi:10.1007/s10508-020-01764-1

Biggs, M. (2021). Revisiting the effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria. *Journal of Pediatric Endocrinology and Metabolism*, *34*, 937–939. doi:10.1515/jpem-2021-0180

Blackwood of North Oxford, B. (2019). *Answer to written question HL15681 asked by Lord Lucas*. UK: House of Lords.

Bouman, M.-B. (2021). Interview. In A. Bakker (Ed.), *The Dutch approach: Fifty years of transgender health care at the VU Amsterdam gender clinic* (pp. 141–146). Los Angeles, CA: Boom.

Bouman, M.-B., van der Sluis, W. B., Buncamper, M. E., Özer, M., Mullender, M. G., & Meijerink, W. J. H. J. (2016). Primary total laparoscopic sigmoid vaginoplasty in transgender women with penoscrotal hypoplasia: A prospective cohort study of surgical outcomes and follow-up of 42 patients. *Plastic and Reconstructive Surgery*, *138*, 614e–623e. doi:10.1097/PRS.0000000000002549

Bowers, M. (2022, March 21). Teen transitions. In *Trans and Gender Diverse Policies, Care, Practices, and Wellbeing Symposium*, Duke University. https://www.facebook.com/dukesgmhealth/videos/704267637246585/

Bradley, S. J., & Zucker, K. J. (1990). Gender identity disorder and psychosexual problems in children and adolescents. *Canadian Journal of Psychiatry*, *35*, 477–486. doi:10.1177/070674379003500603

Brik, T., Vrouenraets, L. J. J. J., de Vries, M. C., & Hannema, S. E. (2020). Trajectories of adolescents treated with gonadotropin-releasing hormone analogues for gender dysphoria. *Archives of Sexual Behavior*, *49*, 2611–2618. 10.1007/s10508-020-01660-8

Butler, G. (2016, June 19). *How effective is puberty suspension with GnRH analogues*. World Professional Association for Transgender Health. http://wpath2016.conferencespot.org/62620-wpathv2-1.3138789/t001-1.3140111/f001-1.3140333/0706-000441-1.3140337

Butler, G., Adu-Gyamfi, K., Clarkson, K., El Khairi, R., Kleczewski, S., Roberts, A., … Carmichael, P. (2022). Discharge outcome analysis of 1089 transgender young people referred to paediatric endocrine clinics in England 2008–2021. *Archives of Disease in Childhood*. doi:10.1136/archdischild-2022-324302

Butler, G., De Graaf, N., Wren, B., & Carmichael, P. (2018). Assessment and support of children and adolescents with gender dysphoria. *Archives of Disease in Childhood*, *103*, 631–636. 10.1136/archdischild-2018-314992

Byng, R., Bewley, S., Clifford, D., & McCartney, M. (2018). Gender-questioning children deserve better science. *Lancet*, *392*(10163), 2435. doi:10.1016/S0140-6736(18)32223-2

Carmichael, P. (2016, June 18). *Time to reflect: Gender dysphoria in children and adolescents, defining best practice in a fast changing context*. World Professional Association for Transgender Health. http://av-media.vu.nl/VUMedia/Play/581e58c338984dafb455c72c56c0bfa31d?catalog=2d190891-4e3f-4936-a4fa-2e9766ae0d0d

Carmichael, P., Butler, G., Masic, U., Cole, T. J., De Stavola, B. L., Davidson, S., Skageberg, E. M., Khadr, S., & Viner, R. (2021). Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK. *PLoS One*, *16*, e0243894. 10.1371/journal.pone.0243894

Carmichael, P., Phillott, S., Dunsford, M., Taylor, A., & de Graaf, N. (2016, June 19). *Gender dysphoria in younger children: Support and care in an evolving context*. World Professional Association for Transgender Health. http://wpath2016.conferencespot.org/62620-wpathv2-1.3138789/t001-1.3140111/f009a-1.3140266/0706-000523-1.3140268

Cassio, A., Cacciari, E., Balsamo, A., Bal, M., & Tassinari, D. (1999). Randomised trial of LHRH analogue treatment on final height in girls with onset of puberty aged 7.5–8.5 years. *Archives of Disease in Childhood*, *81*, 329–332. 10.1136/adc.81.4.329

Chen, D., Abrams, M., Clark, L., Ehrensaft, D., Tishelman, A. C., Chan, Y.-M., … Hidalgo, M. A. (2021). Psychosocial characteristics of transgender youth seeking gender-affirming medical treatment: Baseline findings from the Trans Youth Care study. *Journal of Adolescent Health*, *68*, 1104–1111. doi:10.1016/j.jadohealth.2020.07.033

Cohen, S., & Cosgrove, C. (2009). *Normal at any cost: Tall girls, short boys, and the medical industry's quest to manipulate height*. New York, NY: Penguin Publishing Group.

Cohen-Kettenis, P. (2021). Interview. In A. Bakker (Ed.), *The Dutch approach: Fifty years of transgender health care at the VU Amsterdam gender clinic* (pp. 112–118). Los Angeles, CA: Boom.

Cohen-Kettenis, P., van Goozen, S. H. M., & Cohen, L. (1998). Transsexualism during adolescence. In D. Di Ceglie & D. Freedman (Eds.), *A stranger in my own body: Atypical gender identity development and mental health* (pp. 118–125). London: Karnac Books.

Cohen-Kettenis, P. T. (1994). Die Behandlung von Kindern und Jugendlichen mit Geschlechtsidentitätsstörungen an der Universität Utrecht. *Zeitschrift für Sexualforschung*, *7*, 231–239.

Cohen-Kettenis, P. T. (2016, November 25). *Lessons Learned in 10+ Years of Experience Using Puberty Blockers at VUMC Amsterdam*. 10 ans de la Fondation Agnodice, Lausanne. https://vimeo.com/241880094/46dd8a76af

Cohen-Kettenis, P. T., Delemarre-van de Waal, H. A., & Gooren, L. J. G. (2008). The treatment of adolescent transsexuals: Changing insights. *Journal of Sexual Medicine*, *5*, 1892–1897. doi:10.1111/j.1743-6109.2008.00870.x

Cohen-Kettenis, P. T., Dillen, C. M., & Gooren, L. J. (2000). De behandeling van jonge transseksuelen in Nederland. *Nederlands Tijdschrift voor Geneeskunde*, *144*, 698–702.

Cohen-Kettenis, P. T., & Gooren, L. J. G. (1999). Transsexualism: A review of etiology, diagnosis and treatment. *Journal of Psychosomatic Research*, *46*, 315–333.

Cohen-Kettenis, P. T., & Pfäfflin, F. (2003). *Transgenderism and intersexuality in childhood and adolescence: Making choices*. Thousand Oaks, CA: Sage.

Cohen-Kettenis, P. T., Schagen, S. E. E., Steensma, T. D., de Vries, A. L. C., & Delemarre-van de Waal, H. A. (2011). Puberty suppression in a gender-dysphoric adolescent: A 22-year follow-up. *Archives of Sexual Behavior*, *40*, 843–847. doi:10.1007/s10508-011-9758-9

Cohen-Kettenis, P. T., & van Goozen, S. H. M. (1997). Sex reassignment of adolescent transsexuals: A follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *36*, 263–271. doi:10.1097/00004583-199702000-00017

Cohen-Kettenis, P. T., & van Goozen, S. H. M. (1998). Pubertal delay as an aid in diagnosis and treatment of a transsexual adolescent. *European Child and Adolescent Psychiatry*, *7*, 246–248. doi:10.1007/s007870050073

Costa, R., Dunsford, M., Skagerberg, E., Holt, V., Carmichael, P., & Colizzi, M. (2015). Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. *Journal of Sexual Medicine*, *12*, 2206–2214. doi:10.1111/jsm.13034

de Graaf, N. M., Cohen-Kettenis, P. T., Carmichael, P., de Vries, A. L. C., Dhondt, K., Laridaen, J., Pauli, D., Ball, J., & Steensma, T. D. (2018). Psychological functioning in adolescents referred to specialist gender identity clinics across Europe: A clinical comparison study between four clinics. *European Child and Adolescent Psychiatry*, *27*, 909–919. doi:10.1007/s00787-017-1098-4

de Hingh, V. (2021). Interview. In A. Bakker (Ed.), *The Dutch approach: Fifty years of transgender health care at the VU Amsterdam gender clinic* (pp. 112–118). Los Angeles, CA: Boom.

de Vries, A. L. C. (2010). *Gender dysphoria in adolescents: Mental health and treatment evaluation* (PhD Thesis), Vrije Universiteit, Amsterdam. https://research.vu.nl/en/publications/gender-dysphoria-in-adolescents-mental-health-and-treatment-evalu

de Vries, A. L. C. (2022). Ensuring care for transgender people who need it: Response to 'reconsidering informed consent for trans-identified children, adolescents and young adults'. *Journal of Sex and Marital Therapy*. doi:10.1080/0092623X.2022.2084479

de Vries, A. L. C., & Cohen-Kettenis, P. T. (2012). Clinical management of gender dysphoria in children and adolescents: The Dutch approach. *Journal of Homosexuality*, *59*, 301–320. doi:10.1080/00918369.2012.653300

de Vries, A. L. C., Klink, D., & Cohen-Kettenis, P. T. (2016). What the primary care pediatrician needs to know about gender incongruence and gender dysphoria in children and adolescents. *Pediatric Clinics of North America*, *63*, 1121–1135. 10.1016/j.pcl.2016.07.011

de Vries, A. L. C., McGuire, J. K., Steensma, T. D., Wagenaar, E. C. F., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2014). Young adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*, *134*, 696–704. doi:10.1542/peds.2013-2958

de Vries, A. L. C., Steensma, T. D., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2011). Puberty suppression in adolescents with gender identity disorder: A prospective follow-up study. *Journal of Sexual Medicine*, *8*, 2276–2283. doi:10.1111/j.1743-6109.2010.01943.x

Delemarre-van de Waal, H. A. (2014). Early medical intervention in adolescents with gender dysphoria. In B. P. C. Kreukels, T. D. Steensma, & A. L. C. de Vries (Eds.), *Gender dysphoria and disorders of sex development* (pp. 193–203). Berlin: Springer.

Delemarre-van de Waal, H. A., & Cohen-Kettenis, P. T. (2006). Clinical management of gender identity disorder in adolescents: A protocol on psychological and paediatric endocrinology aspects. *European Journal of Endocrinology*, *155*(suppl_1), S131–S137. doi:10.1530/eje.1.02231

Di Ceglie, D. (2018). The use of metaphors in understanding atypical gender identity development and its psychosocial impact. *Journal of Child Psychotherapy*, *44*, 5–28. doi:10.1080/0075417X.2018.1443151

Everaerd, W., Swaab, H., Gooren, L., Megens, J., & van Trotsenburg, M. (2014). Preface. In B. P. C. Kreukels, T. D. Steensma, & A. L. C. De Vries (Eds.), *Gender dysphoria and disorders of sex development: Progress in care and knowledge* (pp. vii–xxi). Berlin: Springer.

FG. (2021). Interview. In A. Bakker (Ed.), *The Dutch approach: Fifty years of transgender health care at the VU Amsterdam gender clinic* (pp. 131–132). Los Angeles, CA: Boom.
Gender Identity Research and Education Society. (2005). *Consensus Report on Symposium in May 2005*. https://www.gires.org.uk/consensus-report-on-symposium-in-may-2005/
Gender Identity Research and Education Society. (2006). *Final Report to the Nuffield Foundation*. https://www.gires.org.uk/gires-final-report-to-the-nuffield-foundation/
Gijs, L., & Gooren, L. (1996). Hormonal and psychopharmacological interventions in the treatment of paraphilias: An update. *Journal of Sex Research*, *33*, 273–290. doi:10.1080/00224499609551845
Gill-Peterson, J. (2018). *Histories of the transgender child*. Minneapolis, MN: University of Minnesota Press.
Glass, K. (2012, January 22). A boy's own story. *Sunday Times*. https://www.thetimes.co.uk/article/a-boys-own-story-2wpctfb6pxt
Gooren, L. J. G. (1993). Closing speech. In Transsexualism, medicine and law: Proceedings of the 23rd Colloquy on European Law (pp. 233–238). Council of Europe Publishing.
Gooren, L. (2021). Interview. In A. Bakker (Ed.), *The Dutch approach: Fifty years of transgender health care at the VU Amsterdam gender clinic*. Los Angeles, CA: Boom.
Gooren, L., & Delemarre-van de Waal, H. (1996). The feasibility of endocrine interventions in juvenile transsexuals. *Journal of Psychology and Human Sexuality*, *8*, 69–74. doi:10.1300/J056v08n04_05
Green, R. (1968). Childhood cross-gender identification. *Journal of Nervous and Mental Disease*, *147*, 500–509. doi:10.1097/00005053-196811000-00006
Green, R. (1987). *The sissy boy syndrome: The development of homosexuality*. London: Yale University Press.
Green, R. (2008, Autumn). A Tale of Two Conferences. *GT News*, 73.
Groskop, V. (2008, August 14). 'My body is wrong'. *Guardian*. https://www.theguardian.com/society/2008/aug/14/children.youngpeople
Harry Benjamin International Gender Dysphoria Association. (1985). Standards of care: The hormonal and surgical sex reassignment of gender dysphoric persons. *Archives of Sexual Behavior*, *14*, 79–90. doi:10.1007/BF01541354
Harry Benjamin International Gender Dysphoria Association. (2001). *Standards of Care for Gender Identity Disorders, Sixth Version*. http://www.genderpsychology.org/transsexual/hbsoc_1990.html
Hartocollis, A. (2015, June 17). The new girl in school: Transgender surgery at 18. *New York Times*. https://www.nytimes.com/2015/06/17/nyregion/transgender-minors-gender-reassignment-surgery.html
Hausman, B. L. (1995). *Changing sex: Transsexuality, technology and the idea of gender*. Durham, NC: Duke University Press.
Hayes, P. (2016). Early puberty, medicalisation and the ideology of normality. *Women's Studies International Forum*, *56*, 9–18. doi:10.1016/j.wsif.2016.01.003
Hayes, P. (2017). Commentary: Cognitive, emotional, and psychosocial functioning of girls treated with pharmacological puberty blockage for idiopathic central precocious puberty. *Frontiers in Psychology*, *8*. doi:10.3389/fpsyg.2017.00044
Health Research Authority. (2019, October 14). *Investigation into the study 'Early pubertal suppression in a carefully selected group of adolescents with gender identity disorders'*. https://www.hra.nhs.uk/about-us/governance/feedback-raising-concerns/investigation-study-early-pubertal-suppression-carefully-selected-group-adolescents-gender-identity-disorders/
Hembree, W. C., Cohen-Kettenis, P., Delemarre-van de Waal, H. A., Gooren, L. J., Meyer, W. J., Spack, N. P., Tangpricha, V., & Montori, V. M. (2009). Endocrine treatment of transsexual persons: An Endocrine Society clinical practice guideline. *Journal of Clinical Endocrinology and Metabolism*, *94*, 3132–3154. doi:10.1210/jc.2009-0345
Horton, C. (2022). "I didn't want him to disappear": Parental decision-making on access to puberty blockers for trans early adolescents. *Journal of Early Adolescence*. doi:10.1177/02724316221107076
Hough, D., Bellingham, M., Haraldsen, I. R. H., McLaughlin, M., Rennie, M., Robinson, J. E., Solbakk, A. K., & Evans, N. P. (2017). Spatial memory is impaired by peripubertal GnRH agonist treatment and testosterone replacement in sheep. *Psychoneuroendocrinology*, *75*, 173–182. doi:10.1016/j.psyneuen.2016.10.016
Hough, D., Bellingham, M., Haraldsen, I. R., McLaughlin, M., Robinson, J. E., Solbakk, A. K., & Evans, N. P. (2017). A reduction in long-term spatial memory persists after discontinuation of peripubertal GnRH agonist treatment in sheep. *Psychoneuroendocrinology*, *77*, 1–8. doi:10.1016/j.psyneuen.2016.11.029
Hsieh, S., & Leininger, J. (2014). Resource list: Clinical care programs for gender-nonconforming children and adolescents. *Pediatric Annals*, *43*, 238–244. 10.3928/00904481-20140522-11
Indremo, M., Jodensvi, A. C., Arinell, H., Isaksson, J., & Papadopoulos, F. C. (2022). Association of media coverage on transgender health with referrals to child and adolescent gender identity clinics in Sweden. *JAMA Network Open*, *5*, e2146531. doi:10.1001/jamanetworkopen.2021.46531
Ioannidis, J. P. A. (2005). Contradicted and initially stronger effects in highly cited clinical research. *Journal of the American Medical Association*, *294*, 218–228. doi:10.1001/jama.294.2.218
Jennings, J., & Jennings, J. (2016). Trans teen shares her story. *Pediatrics in Review*, *37*, 99–100. 10.1542/pir.2016-002
Johnson, R. T., & Dickersin, K. (2007). Publication bias against negative results from clinical trials: Three of the seven deadly sins. *Nature Clinical Practice Neurology*, *3*, 590–591. doi:10.1038/ncpneuro0618

Joseph, T., Ting, J., & Butler, G. (2019). The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: Findings from a large national cohort. *Journal of Pediatric Endocrinology and Metabolism*, *32*, 1077–1081. doi:10.1515/jpem-2019-0046

Keira Bell and Mrs A v Tavistock NHS Trust (2020). Her Majesty's High Court of Justice in England. Case [2020] EWHC 3274 (Admin).

Kidd, K. M., Sequeira, G. M., Douglas, C., Paglisotti, T., Inwards-Breland, D. J., Miller, E., & Coulter, R. W. S. (2021). Prevalence of gender-diverse youth in an urban school district. *Pediatrics*, *147*, e2020049823. doi:10.1542/peds.2020-049823

Kirsch, I. (2019). Placebo effect in the treatment of depression and anxiety. *Frontiers in Psychiatry*, *10*, article 407. 10.3389/fpsyt.2019.00407

Klink, D., Caris, M., Heijboer, A., van Trotsenburg, M., & Rotteveel, J. (2015). Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. *Journal of Clinical Endocrinology and Metabolism*, *100*, E270–E275. 10.1210/jc.2014-2439

Klotz, F. (2022, April 6). The fractious evolution of pediatric transgender medicine. *Undark Magazine*. https://undark.org/2022/04/06/the-evolution-of-pediatric-transgender-medicine/

Kreukels, B. P. C., & Burke, S. M. (2020). Neurobiology of pediatric gender identity. In M. Forcier, G. Van Schalkwyk, & J. L. Turban (Eds.), *Pediatric gender identity* (pp. 47–62). Berlin: Springer International Publishing. doi:10.1007/978-3-030-38909-3_4

Kuiper, B., & Cohen-Kettenis, P. (1988). Sex reassignment surgery: A study of 141 Dutch transsexuals. *Archives of Sexual Behavior*, *17*, 439–457.

Kuper, L. E., Stewart, S., Preston, S., Lau, M., & Lopez, X. (2020). Body dissatisfaction and mental health outcomes of youth on gender-affirming hormone therapy. *Pediatrics*, *145*, e20193006. doi:10.1542/peds.2019-3006

Laidlaw, M. K., Van Meter, Q. L., Hruz, P. W., Van Mol, A., & Malone, W. J. (2019). Letter to the editor: "Endocrine treatment of gender-dysphoric/gender-incongruent persons: An Endocrine Society clinical practice guideline". *Journal of Clinical Endocrinology and Metabolism*, *104*, 686–687. 10.1210/jc.2018-01925

Lee, J. Y., Finlayson, C., Olson-Kennedy, J., Garofalo, R., Chan, Y.-M., Glidden, D. V., & Rosenthal, S. M. (2020). Low bone mineral density in early pubertal transgender/gender diverse youth: Findings from the Trans Youth Care Study. *Journal of the Endocrine Society*, *4*, bvaa065. doi:10.1210/jendso/bvaa065

Levine, S. B., Abbruzzese, E., & Mason, J. M. (2022). Reconsidering informed consent for trans-identified children, adolescents, and young adults. *Journal of Sex and Marital Therapy*. doi:10.1080/0092623X.2022.2046221

Li, G., Kung, K. T. F., & Hines, M. (2017). Childhood gender-typed behavior and adolescent sexual orientation: A longitudinal population-based study. *Developmental Psychology*, *53*, 764–777. doi:10.1037/dev0000281

Lindgren, T. W., & Pauly, I. B. (1975). A body image scale for evaluating transsexuals. *Archives of Sexual Behavior*, *4*, 639–656. doi:10.1007/BF01544272

Lopez, C. M., Solomon, D., Boulware, S. D., & Christison-Lagay, E. R. (2018). Trends in the use of puberty blockers among transgender children in the United States. *Journal of Pediatric Endocrinology and Metabolism*, *31*, 665–670. doi:10.1515/jpem-2018-0048

Manning, S., & Adams, S. (2014, May 17). NHS to give sex change drugs to nine-year-olds. *Mail on Sunday*. https://www.dailymail.co.uk/news/article-2631472/NHS-sex-change-drugs-nine-year-olds-Clinic-accused-playing-God-treatment-stops-puberty.html

Marumo, K., Baba, S., & Murai, M. (1999). Erectile function and nocturnal penile tumescence in patients with prostate cancer undergoing luteinizing hormone-releasing hormone agonist therapy. *International Journal of Urology*, *6*, 19–23. doi:10.1046/j.1442-2042.1999.06128.x

Medical Practitioners Tribunal Service. (2022). *Record of Determinations: Dr Michael Webberley (2620107)*. https://www.mpts-uk.org/-/media/mpts-rod-files/dr-michael-webberley-25-may-22.pdf

Mermaids. (2007). Obtaining help from the Children's Hospital Boston. In Mermaids *Annual Meeting*. http://www.gires.org.uk/wp-content/uploads/2014/08/mermaids-presentation.ppt

Money, J. (1994). The Concept of gender identity disorder in childhood and adolescence after 39 years. *Journal of Sex and Marital Therapy*, *20*, 163–177. doi:10.1080/00926239408403428

Money, J. (1998). Foreword. In D. Di Ceglie & D. Freedman (Eds.), *A stranger in my own body: Atypical gender identity development and mental health* (pp. xvii–xviii). London: Karnac Books.

Money, J., & Russo, A. J. (1979). Homosexual outcome of discordant gender identity/role in childhood: Longitudinal follow-up. *Journal of Pediatric Psychology*, *4*, 29–41. doi:10.1093/jpepsy/4.1.29

Morse, O. (Director). (1996). The wrong body. In *The Decision*. London: Windfall Films (Channel 4).

Mul, D., Versluis-den Bieman, H., Slijper, F., Oostdijk, W., Waelkens, J., & Drop, S. (2007). Psychological assessments before and after treatment of early puberty in adopted children. *Acta Paediatrica*, *90*, 965–971. doi:10.1111/j.1651-2227.2001.tb01349.x

Nataf, Z. (1999). *Interview*. https://rainbowreeltokyo.com/99/English/interview/zacharynataf.html

Navabi, B., Tang, K., Khatchadourian, K., & Lawson, M. L. (2021). Pubertal suppression, bone mass, and body composition in youth with gender dysphoria. *Pediatrics*, *148*, e2020039339. doi:10.1542/peds.2020-039339

Nederlands Tijdschrift voor Geneeskunde. (1989). Transseksualiteit. *Nederlands Tijdschrift voor Geneeskunde*, *133*, 1475.

Negenborn, V. L., van der Sluis, W. B., Meijerink, W. J. H. J., & Bouman, M.-B. (2017). Lethal necrotizing cellulitis caused by ESBL-producing E. coli after laparoscopic intestinal vaginoplasty. *Journal of Pediatric and Adolescent Gynecology*, *30*, e19–e21. doi:10.1016/j.jpag.2016.09.005

NHS England. (2015). *NHS standard contract for Gender Identity Development Service for children and adolescents*.

Nietsch, H. (Director). (2007). *Valentijn*. Amsterdam: VARA.

Niland, P. (Director). (2014). I Am Leo. In *My Life*. London: CBBC.

Notini, L., Earp, B. D., Gillam, L., McDougall, R. J., Savulescu, J., Telfer, M., & Pang, K. C. (2020). Forever young? The ethics of ongoing puberty suppression for non-binary adults. *Journal of Medical Ethics*, *46*, 743–752. doi:10.1136/medethics-2019-106012

Olson-Kennedy, J., Chan, Y.-M., Garofalo, R., Spack, N., Chen, D., Clark, L., … Rosenthal, S. (2019). Impact of early medical treatment for transgender youth: Protocol for the longitudinal, observational Trans Youth Care Study. *JMIR Research Protocols*, *8*, e14434. doi:10.2196/14434

Pang, K. C., de Graaf, N. M., Chew, D., Hoq, M., Keith, D. R., Carmichael, P., & Steensma, T. D. (2020). Association of media coverage of transgender and gender diverse issues with rates of referral of transgender children and adolescents to specialist gender clinics in the UK and Australia. *JAMA Network Open*, *3*, e2011161. doi:10.1001/jamanetworkopen.2020.11161

Pang, K. C., Notini, L., McDougall, R., Gillam, L., Savulescu, J., Wilkinson, D., … Lantos, J. D. (2020). Long-term puberty suppression for a nonbinary teenager. *Pediatrics*, *145*, e20191606. doi:10.1542/peds.2019-1606

Petersen, M. E., & Dickey, R. (1995). Surgical sex reassignment: A comparative survey of International centers. *Archives of Sexual Behavior*, *24*, 135–156. doi:10.1007/BF01541578

Ruttimann, J. (2013, January). Blocking puberty in transgender youth. *Endocrine News*, 16–20.

Schagen, S. E. E., Wouters, F. M., Cohen-Kettenis, P. T., Gooren, L. J., & Hannema, S. E. (2020). Bone development in transgender adolescents treated with GnRH analogues and subsequent gender-affirming hormones. *Journal of Clinical Endocrinology and Metabolism*, *105*, e4252–e4263. doi:10.1210/clinem/dgaa604

Schneider, C., Cerwenka, S., Nieder, T. O., Briken, P., Cohen-Kettenis, P. T., De Cuypere, G., … Richter-Appelt, H. (2016). Measuring gender dysphoria: A multicenter examination and comparison of the Utrecht gender dysphoria scale and the gender identity/gender dysphoria questionnaire for adolescents and adults. *Archives of Sexual Behavior*, *45*, 551–558. doi:10.1007/s10508-016-0702-x

Schneider, M. A., Spritzer, P. M., Soll, B. M. B., Fontanari, A. M. V., Carneiro, M., Tovar-Moll, F., … Lobato, M. I. R. (2017). Brain maturation, cognition and voice pattern in a gender dysphoria case under pubertal suppression. *Frontiers in Human Neuroscience*, *11*, article 528. doi:10.3389/fnhum.2017.00528

Schroor, E. J., Van Weissenbruch, M. M., & Delemarre-van de Waal, H. A. (1995). Long-term GnRH-agonist treatment does not postpone central development of the GnRH pule generator in girls with idiopathic precocious puberty. *Journal of Clinical Endocrinology and Metabolism*, *80*, 1696–1701. doi:10.1210/jcem.80.5.7745021

Singal, J. (2022, April 6). Researchers found puberty blockers and hormones didn't improve trans kids' mental health at their clinic, then they published a study claiming the opposite. *Singal-Minded*. https://jessesingal.substack.com/p/researchers-found-puberty-blockers

Sloan, J. (2011, October 19). I had sex swap op on my 16th birthday. *Sun*. https://www.thesun.co.uk/fabulous/851138/i-had-sex-swap-op-on-my-16th-birthday/

Smith, Y. L. S., van Goozen, S. H. M., & Cohen-Kettenis, P. T. (2001). Adolescents with gender identity disorder who were accepted or rejected for sex reassignment surgery: A prospective follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *40*, 472–481. doi:10.1097/00004583-200104000-00017

Spack, N. (2008). Foreword. In S. A. Brill & R. Pepper, *The transgender child: A handbook for families and professionals* (pp. ix–xi). Jersey City, NJ: Cleis Press.

Spack, N. P., Edwards-Leeper, L., Feldman, H. A., Leibowitz, S., Mandel, F., Diamond, D. A., & Vance, S. R. (2012). Children and adolescents with gender identity disorder referred to a pediatric medical center. *Pediatrics*, *129*, 418–425. doi:10.1542/peds.2011-0907

Staphorsius, A. S., Kreukels, B. P. C., Cohen-Kettenis, P. T., Veltman, D. J., Burke, S. M., Schagen, S. E. E., … Bakker, J. (2015). Puberty suppression and executive functioning: An fMRI-study in adolescents with gender dysphoria. *Psychoneuroendocrinology*, *56*, 190–199. doi:10.1016/j.psyneuen.2015.03.007

Steensma, T. D., McGuire, J. K., Kreukels, B. P. C., Beekman, A. J., & Cohen-Kettenis, P. T. (2013). Factors associated with desistence and persistence of childhood gender dysphoria: A quantitative follow-up study. *Journal of the American Academy of Child and Adolescent Psychiatry*, *52*, 582–590. doi:10.1016/j.jaac.2013.03.016

Stocks, J. (Director). (2011). *I am Jazz: A family in transition*. Los Angeles, CA: Oprah Winfrey Network.

Stoffers, I. E., de Vries, M. C., & Hannema, S. E. (2019). Physical changes, laboratory parameters, and bone mineral density during testosterone treatment in adolescents with gender dysphoria. *Journal of Sexual Medicine*, *16*, 1459–1468. 10.1016/j.jsxm.2019.06.014

SVT. (2022, February 23). *Uppdrag granskning avslöjar: Flera barn har fått skador i transvården*. SVT Nyheter. https://www.svt.se/nyheter/granskning/ug/uppdrag-granskning-avslojar-flera-barn-har-fatt-skador-i-transvarden

Tishelman, A. C., Kaufman, R., Edwards-Leeper, L., Mandel, F. H., Shumer, D. E., & Spack, N. P. (2015). Serving transgender youth: Challenges, dilemmas, and clinical examples. *Professional Psychology: Research and Practice*, *46*, 37–45. doi:10.1037/a0037490

Tominey, C., & Walsh, J. (2019, March 7). NHS transgender clinic accused of covering up negative impacts of puberty blockers on children by Oxford professor. *Telegraph*. https://www.telegraph.co.uk/news/2019/03/07/nhs-transgender-clinic-accused-covering-negative-impacts-puberty/

Tordoff, D. M., Wanta, J. W., Collin, A., Stepney, C., Inwards-Breland, D. J., & Ahrens, K. (2022). Mental health outcomes in transgender and nonbinary youths receiving gender-affirming care. *JAMA Network Open*, *5*, e220978. 10.1001/jamanetworkopen.2022.0978

van de Grift, T. C., van Gelder, Z. J., Mullender, M. G., Steensma, T. D., de Vries, A. L. C., & Bouman, M.-B. (2020). Timing of puberty suppression and surgical options for transgender youth. *Pediatrics*, *146*, e20193653. doi:10.1542/peds.2019-3653

van der Sluis, W. B., de Nie, I., Steensma, T. D., van Mello, N. M., Lissenberg-Witte, B. I., & Bouman, M.-B. (2021). Surgical and demographic trends in genital gender-affirming surgery in transgender women: 40 years of experience in Amsterdam. *British Journal of Surgery*, *109*, 8–11. doi:10.1093/bjs/znab213

Viner, R., Carmichael, P., Ceglie, D. D., Butler, G., Brain, C., Holt, V., …Skagerberg, E. (2010). *An evaluation of early pubertal suppression in a carefully selected group of adolescents with gender identity disorder (v1.0)*.

Viner, R., Carmichael, P., Ceglie, D. D., Butler, G., Brain, C., Holt, V., …Skagerberg, E. (2012). *An evaluation of early pubertal suppression in a carefully selected group of adolescents with gender identity disorder (v1.2)*.

Vlot, M. C., Klink, D. T., den Heijer, M., Blankenstein, M. A., Rotteveel, J., & Heijboer, A. C. (2017). Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. *Bone*, *95*, 11–19. doi:10.1016/j.bone.2016.11.008

Wiepjes, C. M., Blok, C. J., Staphorsius, A. S., Nota, N. M., Vlot, M. C., Jongh, R. T., & Heijer, M. (2020). Fracture risk in trans women and trans men using long-term gender-affirming hormonal treatment: A nationwide cohort study. *Journal of Bone and Mineral Research*, *35*, 64–70. doi:10.1002/jbmr.3862

Wiepjes, C. M., Nota, N. M., de Blok, C. J. M., Klaver, M., de Vries, A. L. C., Wensing-Kruger, S. A., …den Heijer, M. (2018). The Amsterdam cohort of gender dysphoria study (1972–2015): Trends in prevalence, treatment, and regrets. *Journal of Sexual Medicine*, *15*, 582–590. doi:10.1016/j.jsxm.2018.01.016

Wojniusz, S., Callens, N., Sütterlin, S., Andersson, S., De Schepper, J., Gies, I., …Haraldsen, I. R. (2016). Cognitive, emotional, and psychosocial functioning of girls treated with pharmacological puberty blockage for idiopathic central precocious puberty. *Frontiers in Psychology*, *7*, article 1053. doi:10.3389/fpsyg.2016.01053

Wojniusz, S., Vögele, C., Ropstad, E., Evans, N., Robinson, J., Sütterlin, S., …Haraldsen, I. R. H. (2011). Prepubertal gonadotropin-releasing hormone analog leads to exaggerated behavioral and emotional sex differences in sheep. *Hormones and Behavior*, *59*, 22–27. doi:10.1016/j.yhbeh.2010.09.010

Wren, B. (2020). New way of being a person? *Journal of Medical Ethics*, *46*, 755–756. 10.1136/medethics-2020-106584

Zhou, J.-N., Hofman, M. A., Gooren, L. J. G., & Swaab, D. F. (1995). A sex difference in the human brain and its relation to transsexuality. *Nature*, *378*(6552), 68–70. doi:10.1038/378068a0

Zucker, K. J. (2010). The DSM diagnostic criteria for gender identity disorder in children. *Archives of Sexual Behavior*, *39*, 477–498. doi:10.1007/s10508-009-9540-4

Zuger, B. (1984). Early effeminate behavior in boys: Outcome and significance for homosexuality. *Journal of Nervous and Mental Disease*, *172*, 90–97. doi:10.1097/00005053-198402000-00005

PLOS ONE

RESEARCH ARTICLE

Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK

Polly Carmichael[1]*, Gary Butler[1,2,3], Una Masic[1], Tim J. Cole[3], Bianca L. De Stavola[3], Sarah Davidson[1], Elin M. Skageberg[1], Sophie Khadr[3], Russell M. Viner[3]

1 Gender Identity Development Service (GIDS), Tavistock and Portman NHS Foundation Trust, London, United Kingdom, **2** Paediatric Endocrine Service, University College London Hospitals NHS Foundation Trust, London, United Kingdom, **3** UCL Great Ormond Street Institute of Child Health, University College London, London, United Kingdom

* PCarmichael@tavi-port.nhs.uk

Citation: Carmichael P, Butler G, Masic U, Cole TJ, De Stavola BL, Davidson S, et al. (2021) Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK. PLoS ONE 16(2): e0243894. https://doi.org/10.1371/journal.pone.0243894

Editor: Geilson Lima Santana, University of Sao Paulo Medical School, BRAZIL

Received: February 3, 2020

Accepted: November 29, 2020

Published: February 2, 2021

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0243894

Data Availability Statement: The data underlying this study are available from the UK Data Service (DOI: 10.5255/UKDA-SN-854413).

Abstract

Background

In adolescents with severe and persistent gender dysphoria (GD), gonadotropin releasing hormone analogues (GnRHa) are used from early/middle puberty with the aim of delaying irreversible and unwanted pubertal body changes. Evidence of outcomes of pubertal suppression in GD is limited.

Methods

We undertook an uncontrolled prospective observational study of GnRHa as monotherapy in 44 12–15 year olds with persistent and severe GD. Prespecified analyses were limited to key outcomes: bone mineral content (BMC) and bone mineral density (BMD); Child Behaviour CheckList (CBCL) total t-score; Youth Self-Report (YSR) total t-score; CBCL and YSR self-harm indices; at 12, 24 and 36 months. Semistructured interviews were conducted on GnRHa.

Results

44 patients had data at 12 months follow-up, 24 at 24 months and 14 at 36 months. All had normal karyotype and endocrinology consistent with birth-registered sex. All achieved suppression of gonadotropins by 6 months. At the end of the study one ceased GnRHa and 43 (98%) elected to start cross-sex hormones.

There was no change from baseline in spine BMD at 12 months nor in hip BMD at 24 and 36 months, but at 24 months lumbar spine BMC and BMD were higher than at baseline (BMC +6.0 (95% CI: 4.0, 7.9); BMD +0.05 (0.03, 0.07)). There were no changes from baseline to 12 or 24 months in CBCL or YSR total t-scores or for CBCL or YSR self-harm indices, nor for CBCL total t-score or self-harm index at 36 months. Most participants reported

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

positive or a mixture of positive and negative life changes on GnRHa. Anticipated adverse events were common.

Conclusions

Overall patient experience of changes on GnRHa treatment was positive. We identified no changes in psychological function. Changes in BMD were consistent with suppression of growth. Larger and longer-term prospective studies using a range of designs are needed to more fully quantify the benefits and harms of pubertal suppression in GD.

Introduction

Gender dysphoria (GD) describes the experience of incongruence between an individual's experienced gender and the sex they were assigned at birth. GD [1] in children and young people, also known as Gender Incongruence [2] and previously known as Gender Identity Disorder (GID), is associated with considerable distress or impairment in social, school or other important areas of functioning [3,4]. Interventions include psychosocial support, therapy and medical or surgical interventions to align the body with the identified gender [3,5]. Terminology in this field can be challenging [6]. Here we use birth-registered sex to refer to the sex assigned at birth by clinicians based upon external genitalia [6]. Gender identity refers to a young person's personal sense of their gender. We use the terms 'continuation' and 'discontinuation' to refer to GD across childhood and adolescence.

GD in adolescence is highly likely to continue into adult life where gender dysphoria persists after the onset of puberty [3]. Those with earlier onset or more intense GD and those in whom the development of secondary sexual characteristics in puberty is associated with increasing gender dysphoria or psychological distress are more likely to have persistent GD [3,7]. In adolescents with severe and persistent GD, international [8] and national [9–11] guidelines recommend the use of treatments to suppress the rise in sex hormones (oestradiol or testosterone) in young people during puberty. Gonadotropin releasing hormone analogues (GnRHa) are synthetic peptides that work by stimulating gonadotropin release in a tonic fashion which desensitises the gonadotropin receptors, resulting in reversible suppression of sex hormone production.

In GD, GnRHa can be used from the early/middle stages of puberty with the aim of delaying irreversible and unwanted pubertal body changes and giving young people the opportunity to explore their gender identity during a period when puberty is not advancing [3]. This period also allows clinicians more time to assess the stability of young people's gender identity [6]. Despite this treatment being given in mid-puberty it is also called early puberty suppression, where 'early' refers to earlier than the historic practice of suppression after completion of puberty.

Pubertal suppression is currently practised in the majority of international centres across Europe, the Americas and Australasia, as evidenced by a recently published survey of 25 international centres by the European Society of Paediatric Endocrinology (ESPE) [12]. Pubertal suppression with GnRHa as monotherapy is a time-limited strategy, due to the potential for side effects with long-term use. In the UK, for those commencing under age 15 years, use of GnRHa alone ceases after 16 years when young people face a decision to return to the sex hormones produced by their body or begin cross-sex hormones [5]. There are limited data on the outcomes of pubertal suppression in the treatment of young people with GD [3,13]. A recent

systematic review included data on the physical and mental health outcomes of pubertal suppression using GnRHa in over 500 young people [4]. Longer-term follow-up data on pubertal suppression in GD are limited to individuals from four cohorts [14–19].

In 2011 a study was begun to evaluate the proximal outcomes of mid-pubertal suppression using GnRHa in young people with persistent GD (see http://gids.nhs.uk/our-early-intervention-study). Use in the UK began after mid-pubertal suppression had been incorporated into international guidelines [20] and had become available in the USA [21,22], the Netherlands [15], Australia [23] and a number of European countries. The Gender Identity Development Service (GIDS) at the Tavistock and Portman NHS Foundation Trust, London, is a national service for children and young people with GD, drawing from England, Wales and Ireland. Mid-pubertal suppression was offered by the GIDS from 2011 initially only within an ethically approved uncontrolled observational research study with prospective data collection, where all participants received GnRHa. We anticipated that we would recruit 10–15 young people per year for 3 years and follow them up to the end of monotherapy with GnRHa. At the time, a randomised controlled study was not considered feasible due to very small numbers and inability to retain participants in the control arm, as the control treatment would have resulted in progression into near complete puberty and an increasing number of UK families were accessing mid-pubertal suppression internationally. Allocation blinding was also not considered feasible in young people using a product requiring monthly injections.

Here we describe the short-term outcomes of 44 young people with GD from this research cohort, recruited aged 12–15 years and followed to the end of GnRHa monotherapy after age 16 years. This paper describes their medical, psychological and social outcomes during the GnRHa treatment pathway up to the point of decisions about whether or not to undertake further physical treatment. The aims of the study as defined at inception in 2011 were:

1. To evaluate the benefits and risks for physical and mental health and wellbeing of mid-pubertal suppression in adolescents with GD
2. To add to the evidence base regarding the efficacy of GnRHa treatment for young people with GD
3. To evaluate continuation and discontinuation of GD and the continued wish for gender reassignment within this group.

Methods

We undertook an uncontrolled prospective observational study of GnRHa monotherapy in a highly selected group of young people with persistent and severe GD.

Participants

The cohort consisted of 44 sequentially eligible young people, aged 12 to 15 years, who were recruited between April 2011 and April 2014 and who commenced GnRHa treatment between June 2011 and April 2015. They were all recruited from patients referred to the GIDS.

Eligibility criteria were chosen to match those used for a Netherlands cohort [24], namely that the young person:

A. is aged 12–15 years

B. Psychological criteria

1. has been seen by the GIDS for at least 6 months and attended at least 4 interviews for assessment and therapeutic exploration of their gender identity development.

2. psychological stability sufficient to withstand the stresses of medical treatment for GID.
3. fulfils the following criteria relating to GID:
   a. Throughout childhood (defined as over 5 years) the adolescent has demonstrated an intense pattern of cross-gendered behaviours and cross-gender identity.
   b. The adolescent has gender dysphoria that is significantly increased with the onset of puberty. Following assessment the clinician(s) working with the young person deem that there is a high likelihood of the young person experiencing severe psychological distress consequent on experiencing full pubertal development before pubertal suppression is implemented.
4. The young person and their parents/guardians are actively requesting pubertal suppression.
5. is able to give informed consent.

C. Physical/medical criteria

1. is in established puberty:
   - For birth-registered males Tanner (genital and pubic hair (PH)) stage 3 and above.
   - For birth-registered females Tanner (breast and PH) stage 2 and above.
     The rationale for the sex difference was that the pubertal growth spurt which early intervention aims to avoid occurs typically two years earlier in females (Tanner stage 2–3) than in males (Tanner stage 3–4), thus earlier intervention is required in females.
2. has normal endocrine function and karyotype consistent with birth registered sex.

Note that the presence of mildly elevated androgens in birth registered females consistent with polycystic ovarian syndrome is not an exclusion criterion.

Exclusion criteria:

1. Inability to participate with full investigatory protocol e.g. needle phobia, failure to attend for tests and scans.
2. Body mass index (BMI) <2nd centile for age and birth-registered sex [20].
3. Serious psychiatric conditions (e.g. psychosis, bipolar condition, anorexia nervosa, severe body-dysmorphic disorder unrelated to GD).
4. Inability to give informed consent according to the Fraser/Gillick guidelines.
5. Low spine or hip bone mineral density (BMD) on DXA scan: more than 2 SD below expected BMD for age and birth-registered sex. In exceptional circumstances a low BMD was acceptable if:
   i. it was felt to be clinically appropriate by the treating clinicians, who felt that on the balance of risks, pubertal suppression was justified despite the later risk of osteoporosis
   ii. the young person and parents understood the risks of GnRHa treatment for bone density (i.e. potential risks of later osteoporosis)
   iii. The young person and parents consented to more frequent monitoring of BMD (repeat DXA scans 6 months after starting GnRHa and yearly thereafter while on GnRHa) despite the small DXA radiation dose

iv. The young person and parents consented to stopping treatment if raw BMD fell whilst on GnRHa.

The treatment

The treatment under study was suppression of puberty using the GnRHa *triptorelin* together with psychosocial support and therapy, from study entry until the end of the GnRHa monotherapy pathway at age 16 years or older. GnRHa monotherapy ceased when young people either started cross-sex hormones (and continued on GnRHa) or stopped GnRHa. Treatment duration was therefore from 1 to 4 or 5 years depending on age at study entry. Consenting young people were given triptorelin 3.75mg by intramuscular injection every 28 days during the treatment period. Two participants who found monthly injections difficult were moved to a ten-weekly preparation of 11.25mg of triptorelin. The aim of treatment was to suppress gonadotropins and sex hormones to near pre-pubertal levels [13]. Continued regular attendance for psychological support and therapy throughout the study was a precondition of GnRHa prescription. In addition local psychological services provided support for co-occuring difficulties for participants as required.

Procedures and pathway

All young people and families attending the GIDS during the study period were provided with an information leaflet about research underway within the unit. Those wishing to find out more about the study discussed it with their GIDS clinicians and those deemed likely to be eligible were given detailed written study information. Those wanting to participate were invited to a medical clinic at UCLH for an initial discussion. At the first medical clinic, young people and families were seen by a senior paediatric endocrinology clinician together with a senior GIDS clinician, who discussed with the family the then current state of knowledge and rationale for treatment, eligibility criteria and potential risks and benefits of participation. Risks included the anticipated side-effects of GnRHa treatment including symptoms resulting from the withdrawal of sex steroids (headaches, hot flushes), fatigue, loss of libido and low mood, the potential that treatment could influence the continuation of their GD and the potential for unknown risks. It was emphasised that young people needed to continue with both regular medical and psychosocial follow-up during the study and that treatment would cease if they did not comply with the treatment or monitoring requirements. A full medical history was elicited and the clinicians also reviewed a summary of the psychological history and assessment from the GIDS. In this visit information sheets were re-provided if families had lost them or forgotten details of the study. If young people and families remained interested in participation, medical investigations were organised and families were invited for a repeat discussion and a formal evaluation of eligibility at a second medical clinic visit approximately 3 months later. Families were asked to think about the issues raised in the meeting and to discuss with their GIDS clinicians if necessary, in order to discuss further at the second visit.

At the second medical clinic visit, the same clinicians repeated the discussion of risks and benefits and explored understanding with the young person and family. A chaperoned medical examination was undertaken including pubertal assessment and the results of medical investigations were reviewed. Endocrine and GIDS clinicians jointly reviewed eligibility and offered participation in the study to those deemed eligible.

The implications of treatment for fertility were discussed at the first and second medical visits and all young people were urged to consider storing gametes before starting GnRHa. Access to storage depended on regional availability within the NHS. Note that counselling on fertility

continued across the study, and clinicians periodically checked with young people who had decided against storage whether they wished to revisit their decision.

Informed consent was obtained in writing from both the young person and a parent or carer holding parental responsibility. The ability of the young person and parents to give informed consent was assessed jointly by the senior adolescent endocrine and GIDS clinicians, informed by written notes from the GIDS team. The consent forms were read with the young person and the parent by the clinicians to be sure they fully understood the information on the forms before signing.

48 young people and families attended the medical clinics for discussion of participation in the trial, of whom 44 wished to participate. Eight young people (7 birth assigned males) were not eligible for participation at the second medical visit as they were not yet sufficiently advanced in puberty. They were followed up every 3–6 months and entered the study subsequently when sufficiently advanced in puberty (median waiting time 7 months).

The date of signing the consent form was taken as the start of study treatment, although it frequently took one to three months for GnRHa treatment to start due to administrative requirements. Participants were followed up in the endocrine clinic, 3–6 monthly in the first 18 months and 12-monthly thereafter, till the end of the treatment pathway, defined as the date on or after the 16th birthday when a decision was made to either cease GnRHa or start cross-sex hormones. The final participant completed the pathway in February 2019.

Outcomes

The following data were collected:

A. Baseline explanatory variables

1. Sex and gender: Young people were classified by their sex assigned at birth (birth-registered sex) and self-identified gender.
2. Ethnicity: Ethnicity was obtained from clinic records. For analysis, ethnicity was grouped as white, South Asian, black or mixed.
3. Puberty: Pubertal status at baseline was classified using information on genital/breast and pubic hair Tanner stages as appropriate. This was summarized into a single pubertal stage, with the breast/genital stage taking precedence if there was discrepancy between breast/genital and public hair stage.
4. Clinical data: These consisted of a) identification of normal phenotype on physical examination for birth-registered sex; b) venepuncture assessment of endocrinology (gonadotropins, prolactin, oestrogen or testosterone, adrenal androgens, thyroid function; and a short synacthen test in birth-registered females only), karyotype, full blood count, renal and liver function, calcium and vitamin D; and c) imaging including wrist bone age and (in birth-registered females only) pelvic ultrasound scan. Medical assessment at baseline and follow-up was consistent with Endocrine Society guidelines [8,20].

B. Study outcomes

Study outcomes concerned domains including response to treatment, bone health, safety indicators and adverse events, psychological function; participant experience and satisfaction; and decisions regarding treatment following GnRHa. Outcome data were collected at routine clinic visits to GIDS or medical clinics at UCLH and timings therefore varied. For the purposes of these analyses, data for each participant were assigned to baseline (before treatment) and to the closest of the following outcome periods: 12, 24, 36 and 48 months on treatment. For safety and response to pubertal suppression outcomes, data were also examined at 6 months.

1. Response to pubertal suppression

Gonadotropins (LH, FSH), testosterone (in birth-registered males) and oestrogen (birth-registered females) were measured after venepuncture. Height, weight and blood pressure were recorded by trained clinic staff. BMI z-score for age and birth-registered sex was calculated [25]. Menarcheal status and presence/absence of menstrual periods was obtained by report from birth-registered females.

2. Bone health

Bone mineral content (BMC) and bone mineral density (BMD) in the lumbar (L1 to L4) spine and hip (total hip) were measured by dual energy X-ray absorptiometry (DEXA) scans using a Hologic Discovery QDR series model 010–1549 (Hologic Inc, Bedford, MA, USA). BMD z-scores for age and birth-registered sex appropriate to this machine were calculated [26]. BMD z-scores for spine and hip were further adjusted for height (height-adjusted z-scores) using published formulae [27].

3. Safety indicators and adverse events

Blood samples were collected by venepuncture for liver and renal function, full blood count, calcium and vitamin D, prolactin, adrenal androgens and thyroid function. Participants were routinely questioned about adverse events at medical clinic visits, including anticipated events such as headaches, hot flushes or fatigue plus any other unanticipated events.

4. Psychological function

Psychological outcomes included a clinical outcome routinely collected after GIDS appointments and a range of outcomes assessed using questionnaires. A standardised set of psychological questionnaires used in the GIDS clinic was completed at the time young people were deemed potentially eligible and referred to the medical clinic. Questionnaires were completed at home by the young person and parent between GIDS clinical meetings, and a research assistant followed up families to ensure their completion. Questionnaires were repeated approximately every 12 months on treatment.

i. General psychological functioning

The Child Behaviour Checklist (CBCL) (parent report) and Youth Self Report (YSR) (self-report) are general measures of psychological functioning and part of the Achenbach System of Empirically Based Assessment (ASEBA; www.aseba.org). The CBCL consists of 113 questions and is validated for children aged 6–18 years in international population samples [28]. The YSR consists of 112 questions and is validated in international populations of young people aged 11–18 years [29]. Questions in both are scored on a three-point Likert scale (0 = absent, 1 = occurs sometimes, 2 = occurs often), with the time frame for item responses being the past six months. Scoring for both instruments provides a total problems score, an internalizing problems score (items which assess anxious/depressed, withdrawn-depressed, and somatic complaints) and an externalizing score (focusing on rule-breaking and aggressive behaviours). Each questionnaire was scored with Assessment Data Manager Software using ASEBA standard norms and t-scores were generated based on reference data for birth-registered sex and broad age-ranges (here 12–18 years). Higher scores indicate greater morbidity. To account for normative change within our age-range, we used international reference data [29] to transform YSR raw scores into z-scores for year of age. As reference data from the UK were not available, reference data from both Australia and the Netherlands were used.

ii. Self-harm index

Self-harm actions and thoughts were assessed through two questions in each of the CBCL (parent report) and YSR (self-report): Item 18 (I deliberately try to hurt or kill myself) and Item 91 (I think about killing myself). Possible responses for each question were 0 = not true, 1 = somewhat or sometimes true, or 2 = very true or often true. We followed previous studies in calculating a self-harm index score to avoid multiple statistical comparisons across

correlated categorical-response variables. The index was calculated as the sum of the two items in each scale to create an index from 0 to 4 for each of the CBCL and YSR [30–32], a higher score indicating greater self-harm thoughts and behaviour.

iii. Health related quality of life (HRQoL)

This was assessed through separate young person and parent Kidscreen-52 questionnaires, each consisting of 52 items which assess HRQoL across ten dimensions: physical well-being; psychological well-being; moods and emotions; self-perception; autonomy; relations with parents and home life; social support and peers; school environment; social acceptance (bullying); and financial resources. All items use five-point Likert-style scales to assess either the frequency (never-seldom-sometimes-often-always) of certain behaviours/feelings or the intensity of an attitude (not at all–slightly-moderately-very-extremely). The measure was developed for young people aged 8–18 years, with the recall period of one week. The questionnaires provide scores in the form of continuous t-scores for the ten subscales derived from a multinational European sample [33]. Lower scores indicate lower HRQoL, i.e. greater morbidity.

iv. Body image

The Body Image Scale (BIS) is a self-report measure of 30 items used to assess body image satisfaction or dissatisfaction validated for age 12+. The instrument considers 30 body features which the respondent is asked to rate in terms of satisfaction on a five-point scale (1 = very satisfied, 2 = satisfied, 3 = neutral, 4 = dissatisfied, and 5 = very dissatisfied). The BIS provides a total score in the form of a continuous score for the total scale as well as for three subscales assessing primary sexual characteristics, secondary sexual characteristics and 'neutral' characteristics (i.e. non-sexual characteristics, e.g. nose) [34]. Higher scores represent higher degrees of body dissatisfaction.

v. Gender dysphoria

The Utrecht Gender Dysphoria Scale (UGDS) is a self-report measure used to assess the intensity of GD validated for age 12+. It comprises of 12 statements with agreement on a five-point scale (1 = agree completely, 2 = agree somewhat, 3 = neutral, 4 = disagree somewhat, and 5 = disagree completely). There are separate versions for birth-registered males and females. Items are summed to give a single total score, with higher scores indicating greater GD.

vi. Clinical outcomes

The Children's Global Assessment Scale (CGAS) is a rating of functioning in children and young people aged 6–17 years, extensively used as a routine clinical measure in child and adolescent mental health services in the UK. Treating clinicians assign young people a single score between 1 and 100, based on a clinician's assessment of a range of aspects related to a child's psychological and social functioning, with the time period being the previous month. Higher scores indicate better functioning, with categories ranging from 'extremely impaired' (1–10) to 'doing very well' (91–100) [35].

5. Participant experience and satisfaction with GnRHa

Young people were invited to participate in semi-structured qualitative interviews at 6–15 months and 15–24 months after starting GnRHa. Interviews were conducted in person or by telephone with a research assistant. If young people were unavailable, questions were posted to be completed and returned. The interview consisted of 12 questions related to changes young people had experienced in ten domains since starting on GnRHa: life overall, memory, focus, sense of direction, mood, energy levels, relationships with friends, relationships with family, gender role and sexuality. For each domain, young people were asked first about the general direction of change in that domain (whether changes were positive, neutral, negative or mixed positive and negative) and then asked for examples of changes experienced and why they assigned the chosen change rating. At the end of the interview two further questions were asked about change in any other experiences (i.e. allowing open ended responses) and whether

young people wished to continue on GnRHa treatment. Note there was no interview conducted before young people started GnRHa. Interviews were recorded in contemporaneous written notes by the researcher. The questionnaire is provided in the S1 Appendix.

6. Further treatment decisions

Decisions made at the end of the GnRHa pathway were recorded in terms of which if any further treatment for GD young people chose.

Note that other measures of gender dysphoria (Gender Identity Interview; Recalled Childhood Gender Identity Scale) were specified in our original protocol, however they were discontinued during the study as: a) they were historical instruments with poor construct validity and the binary references to male and female roles were challenging for some participants; and b) repeated questioning about gender dysphoria resulted in some distress to respondents. Our protocol had originally included the ASEBA Teacher Report Form (TRF), however we were unable to obtain data from teachers so this outcome was dropped. The Social Responsiveness Scale (SRS) was a baseline only assessment of autistic traits; these data will be analysed in the future.

Analysis plan

Analyses were conducted according to the Statistical Analysis and Dissemination Plan, lodged with the ethics committee that approved the study before the analysis started (see S2 Appendix: Statistical Analysis Plan). The analysis plan was designed to report data on all outcomes but to minimise the likelihood of chance findings due to the large number of outcomes and small sample size. Sample sizes necessarily varied across follow-up as young people were recruited at different ages (12–15 years) but left the study soon after their 16th birthday. All 44 participants had data at 12 months follow-up. As participants necessarily left the study soon after their 16th birthday, numbers reduced after 12 months follow-up as participants could no longer remain in the study. Note this does not represent drop-out. There were 24 left at 24 months, 14 at 36 months and 4 at 48 months. In view of this, outcome reporting was restricted to change from baseline to 12, 24 and 36 months. We made no attempt to account for missing data due to the small sample size and the likelihood of the data missing not at random.

We restricted analyses to primarily descriptive statistics, with formal statistical testing of change across the study restricted to six pre-specified outcomes, i.e.:

1. Overall psychological functioning
    a. parent report: CBCL total t-score
    b. young person self-report: YSR total t-score
2. Self-harm index
    a. parent report: CBCL self-harm index
    b. young person self-report: YSR self-harm index
3. Bone health
    a. BMD and BMC for lumbar spine
    b. BMD and BMC for hip

Assessment of change was through paired t-tests for normally distributed data and the Wilcoxon matched-pairs sign-rank test for non-normal data. The number of formal statistical tests conducted in the study was 16; with overall significance at $p = 0.05$ and a Bonferroni correction, the appropriate threshold for statistical significance is about $p = 0.003$.

In our results and conclusions we refer to change in outcomes only for those that were formally tested. Reporting for other continuous outcomes was restricted to mean and 95% confidence intervals (95%CI) or median and interquartile range (IQR). For categorical outcomes, simple proportions were reported. We reported laboratory tests as normal or abnormal based upon laboratory reference data for age, with the exception of gonadotropins. We did not report data where the sample size was less than 8.

Analysis of potential predictors of outcome was confined a priori to two factors, birth-registered sex and pubertal stage at baseline. Three pre-specified continuous outcomes were examined at 12 months, namely:

1. BMD for lumbar spine
2. YSR total t-score
3. CGAS score

Associations were examined using linear regression of follow-up score on baseline score, adding each baseline factor separately to the model and considering the interaction of predictor with baseline score. All analyses were conducted using Stata 16 (Statacorp, College Station TX).

Responses to the semi-structured interview questionnaires were analysed simply for thematic content in terms of the direction and amount of change that young people experienced in each domain. This involved coding responses about experiences since starting GnRHa into categories; i.e. either positive/improving, negative/deteriorating, both positive and negative, no change or not known. The question on change in sexuality was coded as yes change, no change or not known. Wishes to continue with GnRHa were coded as yes, no or don't know.

To compare our findings with the literature, we drew upon recent reviews [3,4,6,13] and updated a recent review [4] from 1 June 2017 to 31 December 2019 using the same search terms in Medline (see S1 Appendix).

Ethics

Ethical approval for the study was obtained from the National Research Ethics Service (NRES: reference 10/H0713/79) in February 2011. Study consent allowed the use of routinely collected clinical data (medical and psychological) as part of clinical treatment for the study. Study procedures including consent were reviewed by the UK Health Research Authority.

Data sharing. These are highly sensitive data from a small group of vulnerable young people treated in a single service and the risk of identification and disclosure is high. Research ethics permissions at the time the study was undertaken did not include permission to share data. After discussions with the Health Research Authority, UK, an anonymised dataset modified to remove sensitive data and minimise disclosure risk of personal information has been deposited with the UK Data Service.

Results

Participants received psychosocial assessment and support within the GIDS before entering the study for a median of 2.0 years (IQR 1.4 to 3.2; range 0.7 to 6.6). The median time between first medical assessment at UCLH and starting treatment was 3.9 months (IQR 3.0 to 8.4; range 1.6 to 25.7). Median time in the study was 31 months (IQR 20 to 42, range 12 to 59).

Baseline characteristics of the participants by birth-registered sex are shown in Table 1. Median age at consent was 13.6 years (IQR 12.8 to 14.6, range 12.0 to 15.3). A total of 25 (57%) were birth-registered as male and 19 (43%) as female. At study entry, birth-registered males

Table 1. Participant characteristics at baseline.

		Total sample	Birth-registered sex	
		n = 44	male	female
			n = 25	n = 19
Age at consent (years)	Median (IQR)	13.6 (12.8, 14.6)	13.4 (12.7, 14.1)	13.9 (13.5, 14.7)
Ethnic group n (%)	white	39 (89)	24 (96)	15 (79)
	South Asian	1 (2)	1 (4)	0
	black	2 (5)	0	2 (11)
	Mixed ethnicity	2 (5)	0	2 (11)
Pubertal status n (%)	Stage 2	0	0	0
	Stage 3	19 (43)	17 (68)	2 (10)
	Stage 4	16 (36)	5 (20)	11 (58)
	Stage 5	9 (21)	3 (12)	6 (32)
Menarcheal status n (%)	Premenarcheal	-	-	4 (21)
	Post-menarcheal	-	-	15 (79)
Time in study (months)	Median (IQR)	31 (20, 42)	37 (24, 43)	29 (17, 36)
Age at end of pathway (years)	Median (IQR)	16.1 (16.0, 16.4)	16.1 (16.0, 16.5)	16.1 (16.0, 16.3)

At baseline, all participants had normal endocrinology, karyotype, imaging and clinical phenotype on physical examination for birth-registered sex and normal full blood count and liver and renal function. No participants had evidence of disorders of sexual differentiation. Eight participants (18%) had vitamin D insufficiency at baseline and were given vitamin D supplements.

https://doi.org/10.1371/journal.pone.0243894.t001

were predominantly in stage 3 puberty (68%) whilst birth-registered females were predominantly in stages 4 (58%) or 5 (32%) with 79% (15/19) post-menarcheal. 89% of participants were of white ethnicity. Birth-registered females were on average 6 months older than birth-registered males at study entry.

Response to treatment

All participants achieved adequate suppression of gonadotropins and sex hormones by 6 months (mean LH 0.5IU/L; mean FSH 1.4IU/L) and maintained it throughout the study (see Table 2). Liver function, basic haematology and biochemistry were normal in all participants at 3–6 months. All post-menarcheal birth-registered females reported amenorrhoea in the 3 months after starting GnRHa treatment and remained so throughout treatment. No participants reported progression in pubertal development. Height and weight were normal at baseline. Height growth continued through the study but more slowly than expected for age, thus

Table 2. Growth and gonadotropin levels at baseline, 12, 24 and 36 months.

Growth		Baseline		12 months		24 months		36 months	
		n	Mean (95% CI)	n	Mean (95% CI)	n	Mean (95% CI)	n	Mean (95% CI)
Height	z-score	44	0.4 (0.1, 0.7)	44	0.2 (-0.1, 0.4)	24	0.0 (-0.4, 0.4)	14	0.0 (-0.5, 0.5)
Weight	z-score	44	0.8 (0.4, 1.3)	44	0.8 (0.3, 1.3)	24	0.6 (-0.1 1.3)	14	1.0 (0.1, 1.9)
BMI	z-score	44	0.7 (0.2, 1.1)	44	0.7 (0.2, 1.2)	24	0.6 (-0.1, 1.3)	14	1.1 (0.3, 1.9)
Gonadotropins									
LH	IU/L	42*	4.2 (2.8, 5.6)	44	0.60 (0.42, 0.68)	17	0.40 (0.22, 0.60)	7	0.30 (0.14, 0.46)
FSH	IU/L	42*	3.9 (3.2, 4.5)	44	1.3 (1.0, 1.7)	17	1.0 (0.6, 1.5)	7	1.4 (0.7, 2.2)

*In two participants data recorded as normal at baseline were not available.

https://doi.org/10.1371/journal.pone.0243894.t002

height z-score fell over time (Table 2). Weight and BMI z-scores were stable from baseline to 24 months but increased at 36 months.

Three participants had brief periods off GnRHa prior to their 16th birthday. In one, treatment was withdrawn by clinicians due to non-attendance at clinics and restarted 4 months later. Another requested a period off GnRHa to think further about treatment in view of other things happening in their life; they restarted 4 months later. A third, birth-registered male, stopped GnRHa for 9 months to attempt to store sperm, contrary to their earlier decision not to, and restarted afterwards.

Median age at the end of the GnRHa pathway was 16.1 years (Table 1). A quarter of participants made their decision more than six months later, either because they wished to delay due to school exams or other events or because clinicians felt they were not yet ready to make the decision. One young person decided to stop GnRHa and not start cross-sex hormones, due to continued uncertainty and some concerns about side-effects of cross-sex hormones. The remaining 43 (98%) elected to start cross-sex hormones.

Bone mineral density. BMD was available on 44 participants at baseline, 43 at 12 months, 24 at 24 months and 12 at 36 months (Table 3). Numbers were lower for hip than for spine as some hip scans were not done for technical reasons. The table shows mean values at baseline and 12, 24 and 36 months, along with mean baseline values corresponding to the paired samples at each time point. There was no change from baseline in spine or hip at 12 months nor in hip at 24 and 36 months, but at 24 months lumbar spine BMC and BMD were higher than at baseline, as was lumbar BMC at 36 months. Lumbar and hip BMD age-adjusted z-scores were in the normal range at baseline but point-estimates fell at 12 and 24 months but not at 36 months. Point-estimates for height-adjusted z-scores for lumbar and hip BMD also fell at 12 and 24 months but not at 36 months.

Psychological outcomes. For the standardised questionnaires, baseline assessments were conducted at a median of 0.5 (IQR 0.4, 0.8) years before starting treatment, and were available for all 44 participants by self-report and 43 by parental report. Data on the CBCL, YSR, Kidscreen-52, BIS and CGAS were normally distributed whilst those for UGDS and the CBCL and YSR self-harm indices were skewed.

The first psychological follow-up was at a median of 13 (IQR 12, 14) months after start of treatment, with ASEBA data available for 41 participants (parent and self-report). ASEBA data at 24 months (median 25 (21, 28)) were available on 20 young people by parent report and 15 by self-report, and at 36 months (median 36 (29, 39)) on 11 by parent report and 6 by self-report.

Formal testing was undertaken only for key ASEBA outcomes (Table 4). For the CBCL total t-scores, there was no change from baseline to 12, 24 or 36 months. Similarly for the YSR total t-score, there was no change from baseline to 12 or 24 months; YSR data at 36 months (n = 6) were not analysed. There were no significant changes in parent-report CBCL self-harm index scores from baseline to 12, 24 or 36 months, nor for self-report YSR self-harm index scores.

Other psychological outcomes are described in Table 5. Point-estimates of scores on the Kidscreen-52, BIS, UGDS and CGAS showed little change over time."

The pre-specified outcomes of BMD at lumbar spine, YSR total t-score and CGAS score at 12 months, adjusted separately for birth-registered sex and baseline pubertal status, along with the baseline level of the outcome, are shown in Table 6. None of the outcomes were associated with birth-registered sex or pubertal status, and there were no important interactions.

Participant experience, satisfaction and side effects. 41 participants completed interviews at 6–15 months (median 9) and 29 at 15–24 months (median 21); 3 missed both. Fig 1 shows proportions with positive or negative changes for life overall, mood and friendships, with summary data for all questions shown in S1 Appendix (S1 and S2 Tables).

Table 3. Bone mineral density outcomes at baseline, 12, 24 and 36 months.

				12 months					24 months				
		Baseline			Baseline for those followed up	Follow-up	Change	p		Baseline for those followed up	Follow-up	Change	p
		n	Mean (95% CI)	n	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)		n	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)	
Lumbar	BMC	44	39.5 (35.9, 43.1)	42	39.6 (35.8, 43.4)	41.2 (38.2, 44.2)	1.6 (0.2, 3.1)	0.03	24	34.1 (30.3, 37.9)	40.1 (36.7, 43.5)	6.0 (4.0, 7.9)	<0.0001
	BMD	44	0.76 (0.71, 0.80)	43	0.76 (0.71, 0.80)	0.77 (0.72, 0.81)	0.01 (-0.00, 0.03)	0.17	24	0.68 (0.63, 0.74)	0.73 (0.68, 0.78)	0.05 (0.03, 0.07)	0.0001
Hip	BMC	43	25.2 (23.2, 27.1)	39	25.5 (23.4, 27.6)	26.1 (24.4, 27.9)	0.7 (-0.2, 1.5)	0.13	22	23.9 (21.2, 26.6)	26.3 (24.1, 28.6)	2.4 (0.7, 4.1)	0.008
	BMD	43	0.80 (0.75, 0.86)	39	0.81 (0.75, 0.87)	0.82 (0.78, 0.86)	0.01 (-0.02, 0.05)	0.6	22	0.76 (0.68, 0.85)	0.79 (0.74, 0.84)	0.03 (-0.04, 0.10)	0.4
BMD z-scores	Spine	44	-0.3 (-0.7, 0.0)	43	-0.3 (-0.7, 0.1)	-1.0 (-1.3, -0.7)			24	-0.5 (-1.1, 0.0)	-1.5 (-2.1, -0.8)		
	HAZ spine	44	-0.5(-0.8, -0.1)	43	-0.4 (-0.8, -0.1)	-1.0 (-1.3, -0.6)			24	-0.7 (-1.2, -0.1)	-1.3 (-1.9, -0.7)		
	Hip	43	-0.5 (-0.9, -0.1)	39	-0.5 (-0.9, -0.1)	-1.0 (-1.3, -0.6)			21	-0.5 (-1.1, 0.1)	-1.4 (-2.0, -0.9)		
	HAZ hip	43	-0.7 (-1.0, -0.3)	39	-0.6 (-1.0, -0.2)	-0.9 (-1.3, -0.5)			21	-0.5 (-1.1, 0.1)	-1.2 (-1.7, -0.6)		
				36 months									
					Baseline for those followed up	Follow-up	Change	p					
				n	Mean (95% CI)	Mean (95% CI)	Mean (95% CI)						
Lumbar	BMC			12	37.05 (31.0, 43.1)	42.4 (37.4, 47.4)	5.3 (2.8, 7.8)	0.0007					
	BMD			12	0.72 (0.65, 0.80)	0.76 (0.70, 0.82)	0.03 (.00, 0.07)	0.05					
Hip	BMC			12	26.1 (22.1, 30.0)	26.8 (21.2, 32.3)	0.7 (-3.8, 5.2)	0.7					
	BMD			12	(0.82, 0.73, 0.91)	0.81 (0.74, 0.88)	-0.009 (-0.05, 0.03)	0.6					
BMD z-scores	Spine			12	-0.2 (-1.0, 0.6)	-1.5 (-2.2, -0.8)							
	HAZ spine			12	-0.4 (-1.2, 0.3)	-1.3 (-2.2, -0.5)							
	Hip			12	-0.3 (-1.3, 0.6)	-1.1 (-1.8, -0.5)							
	HAZ hip			12	-0.5 (-1.5, 0.5)	-1.0 (-1.8, -0.2)							

BMD: bone mineral density; BMC bone mineral content; HAZ height adjusted z-score.

BMD z-scores were not formally tested–see Methods.

https://doi.org/10.1371/journal.pone.0243894.t003

Most participants reported positive or a mix of positive-negative changes in their life at both time points. At 6–15 months 46% reported only positive changes, including feeling happier, relieved, less facial hair or stopping periods. A further 37% reported both positive and negative changes such as feeling happier but also experiencing hot flushes and headaches. In addition 12% reported overall negative changes namely hot flushes, tiredness, and feeling more emotional, while 5% reported no change. At 15–24 months, 55% reported solely positive changes such as feeling happier, no longer experiencing side effects and feeling more

Table 4. ASEBA outcomes at baseline, 12, 24 and 36 months.

				12 months					24 months				
			Baseline		Baseline for those followed up	Follow-up	Change	p		Baseline for those followed up	Follow-up	Change	p
		n	mean (95% CI)	n	mean (95% CI)	mean (95% CI)	mean (95% CI)		n	mean (95% CI)	mean (95% CI)	mean (95% CI)	
Parent report CBCL	Total problems t-score	43	61.6(58.4, 64.7)	41	61.5(58.2, 64.7)	61.8(58.4, 65.1)	0.3(-2.0, 2.6)	0.8	20	61.2(56.5, 65.8)	60.2(54.6, 65.8)	-1.0(-4.0, 2.1)	0.5
	Externalising problems t-score	43	55.8(52.4, 59.3)	41	55.7(52.1, 59.3)	55.4(51.8, 59.0)			20	55.4(49.9, 60.9)	55.2(48.9, 61.5)		
	Internalising problems t-score	43	62.1(58.7, 65.5)	41	61.8(58.3, 65.3)	62.9(59.5, 66.3)			20	60.4(55.7, 65.1)	60.1(54.6, 65.6)		
Self-report YSR	Total problems t-score	44	57.9(55.0, 60.8)	41	57.6(54.5, 60.6)	58.4(54.6, 62.2)	0.8(-3.1, 4.8)	0.7	15	55.1(50.9, 59.2)	56.5(50.6, 62.5)	1.5(-3.4, 6.3)	0.5
	Total problems z-score (ref: Netherlands)	44	1.01(0.67, 1.36)	41	0.97(0.62, 1.33)	0.99(0.55, 1.42)			15	0.66(0.17,1.15)	0.65(-0.05, 1.36)		
	Total problems z-score (ref: Australia)	44	0.72(0.37, 1.06)	41	0.68(0.32, 1.03)	0.68(0.24, 1.12)			15	0.39(-0.11,0.89)	0.37(-0.32, 1.07)		
	Externalising problems t-score	44	52.3(49.2, 55.5)	41	52.3(49.2, 55.4)	52.5(48.7, 56.3)			15	53.1(48.5, 57.6)	52.3(45.3, 59.4)		
	Internalising problems t-score	44	58.0(54.9, 61.2)	41	57.7(54.3, 61.0)	60.1(55.9, 64.3)			15	53.9(49.9, 58.0)	55.9(50.8, 61.1)		
Self-harm scores													
Parent report CBCL	Median (IQR)	43	0(0, 1)	40	0(0, 1)	0(0, 1)		0.3	20	0(0, 1)	0(0, 1)		>0.9
Self-report YSR	Median (IQR)	43	0(0, 1)	39	0(0, 1)	0(0, 2)		0.4	15	0(0, 0)	0(0, 0)		0.3
				36 months									
					Baseline for those followed up	Follow-up	Change	**p**					
				n	mean (95% CI)	mean (95% CI)	mean (95% CI)						
Parent report CBCL	Total problems t-score			11	62.4(55.1, 69.6)	61.1(52.3, 69.9)	-1.3(-6.6, 4.0)	0.6					
	Externalising problems t-score			11	56.8(48.0, 65.6)	56.2(48.3, 64.1)							
	Internalising problems t-score			11	60.4(53.5, 67.2)	62.5(53.6, 71.5)							
Self-harm scores													
Parent report CBCL	Median (IQR)			11	0(0, 1)	0(0, 1)		0.8					

https://doi.org/10.1371/journal.pone.0243894.t004

comfortable with puberty suspended. A further 17% reported both positive and negative changes including less body hair but continued growth in height, or having clearer skin but also experiencing more hunger, weight gain and tiredness. 17% reported largely negative changes such as mood swings, tiredness and hot flushes whilst 10% reported no change.

Reports of change in mood were mixed. At 6–15 months, the majority reported mood to be improved (49%), mixed changes (such as both feeling happier but experiencing some mood swings; 15%) or no change (7%), however 24% reported negative changes in mood such as

Table 5. Other psychological outcomes at baseline, 12, 24 and 36 months.

		Baseline		12 months		24 months		36 months	
		n	mean (95% CI)	n	mean (95% CI)	n	mean (95% CI)	n	mean (95% CI)
Kidscreen-52 HRQOL									
Parent report CBCL t-scores	Physical wellbeing	42	44.9(41.4, 48.5)	36	40.4(37.5, 43.3)	14	40.5(36.8, 44.2)		
	Psychological Wellbeing	41	39.8(36.7, 42.8)	36	39.0(35.4, 42.6)	14	42.4(36.9, 48)		
	Moods and Emotions	41	40.6(37.6, 43.6)	36	41.2(37.3, 45.1)	14	42.5(36.3, 48.7)		
	Self-perception	42	34.6(32.6, 36.5)	36	34.8(32.0, 37.5)	14	34.8(31.3, 38.2)		
	Autonomy	42	46.2(43.2, 49.2)	36	48.2(45.0, 51.4)	14	46.7(41, 52.4)		
	Parent relations and home life	42	48.1(44.5, 51.6)	35	46.7(42.9, 50.5)	14	49.5(44.1, 54.9)		
	Social support and peers	39	48.0(44.7, 51.4)	36	51.9(48.4, 55.3)	13	51.4(45.6, 57.2)		
	School environment	42	38.2(35.0, 41.4)	35	39.4(35.3, 43.4)	13	43.7(36, 51.3)		
	Social acceptance	39	44.7(40.7, 48.7)	32	42.3(38.1, 46.4)	13	43.5(35.9, 51.2)		
	Financial resources	42	37.9(33.9, 41.9)	36	35.8(31.5, 40.2)	14	36.3(26.4, 46.3)		
Self-report t-scores	Physical wellbeing	42	45.1(41.8, 48.5)	36	41.5(38.0, 45.0)	13	43.9(38.9, 48.9)		
	Psychological Wellbeing	42	43.0(39.6, 46.4)	36	41.1(37.0, 45.2)	14	51(45.8, 56.2)		
	Moods and Emotions	42	46.3(42.7, 49.9)	36	43.9(40.4, 47.3)	14	50.1(45.5, 54.7)		
	Self-perception	42	38.8(36.7, 40.9)	36	37.9(35.1, 40.6)	14	43.1(39.9, 46.2)		
	Autonomy	42	46.6(43.6, 49.6)	36	46.7(42.9, 50.5)	13	51.9(47.4, 56.4)		
	Parent relations and home life	42	49.7(46.2, 53.2)	36	48.7(45.2, 52.3)	14	58.4(53.3, 63.5)		
	Social support and peers	37	45.6(42.5, 48.7)	35	48.1(44.6, 51.6)	14	49.7(44.3,55.1)		
	School environment	41	45.9(42.3, 49.4)	36	44.7(39.7, 49.7)	14	49(43.6, 54.3)		
	Social acceptance	41	47.4(43.5, 51.3)	33	45.5(40.9, 50.1)	13	53.6(46.3, 60.8)		
	Financial resources	42	42.2(38.1, 46.3)	34	43.2(38.2, 48.1)	14	46.3(39.1, 53.5)		
Body image scale	Overall score	42	3.1(2.8, 3.3)	40	3.2(3.0, 3.4)	16	3(2.7, 3.2)	8	3.1(2.4, 3.7)
	Primary characteristics score	42	4.5(4.2, 4.7)	39	4.3(4.2, 4.5)	16	4.5(4.3, 4.7)	8	4.2(3.9, 4.5)
	Secondary characteristics score	41	2.9(2.6, 3.1)	40	3(2.8, 3.3)	16	2.9(2.5, 3.2)	8	2.9(2, 3.8)
	Neutral characteristics score	42	2.5(2.203, 2.707)	40	2.7(2.5, 3.0)	-	-		
Utrecht Gender dysphoria score	Median (IQR)	41	4.8(4.6, 5.0)	40	4.7(4.6, 5.0)	18	4.7(4.3, 5.0)		
Clinical outcome									
CGAS global score	Mean (95% CI)	42	62.9(59.6, 66.2)	35	64.1(59.9, 68.3)	18	65.7(59.6, 71.8)	12	66.0(58.1, 73.9)

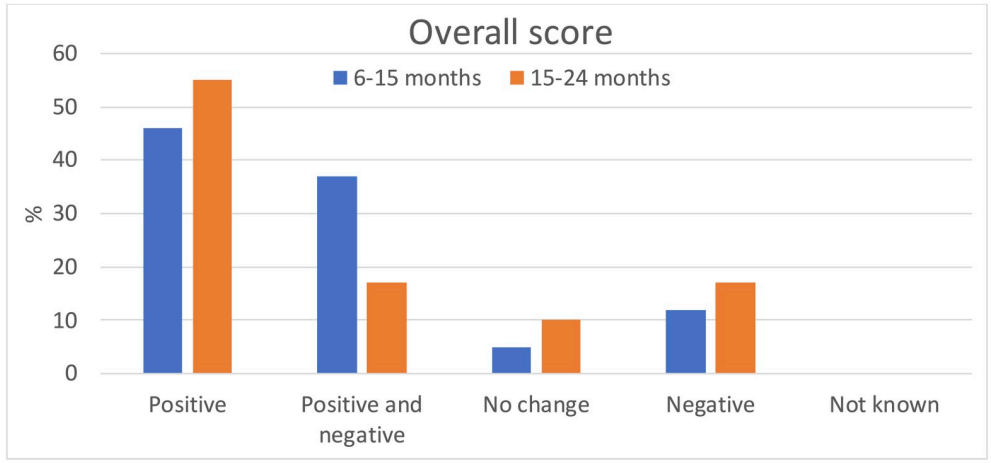
Note: Change in outcomes in this Table were not formally tested.

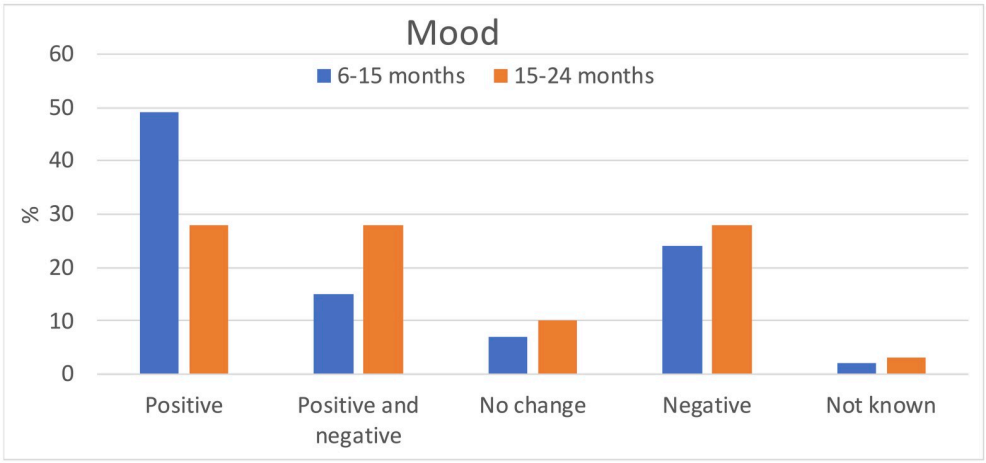
https://doi.org/10.1371/journal.pone.0243894.t005

Table 6. Associations between birth-registered sex and baseline pubertal status and outcomes at 12 months.

		Outcomes at 12 months adjusted for baseline								
		BMD at lumbar spine			YSR total t-score			GCAS score		
		n	Coefficient (95% CI)	p	n	Coefficient (95% CI)	p	n	Coefficient (95% CI)	p
Birth-registered sex										
Main effect (baseline value of outcome)		43	0.86 (0.75, 0.97)	<0.0001	41	0.43 (0.05, 0.82)	0.03	33	0.74 (0.42, 1.06)	<0.0001
Birth-registered sex	Male (ref)		0			0			0	
	Female		-0.02 (-0.05, 0.01)	0.2		2.1 (-5.2, 9.4)	0.6		-3.2 (-10.0, 3.5)	0.3
Pubertal status										
Main effect (baseline value of outcome)		43	0.85 (0.72, 0.97)	<0.0001	41	0.43 (0.01, 0.84)	0.04	33	0.69 (0.37, 1.00)	<0.0001
Pubertal stage at baseline	3		0.008 (-0.03, 0.04)	0.7		0.2 (-8.3, 8.7)	0.9		1.6 (-5.5, 8.8)	0.6
	4 (ref)		0			0			0	
	5		-0.009 (-0.05, 0.03)	0.7		0.4 (-9.9, 10.8)	0.9		-7.9 (-17.6, 1.8)	0.11

https://doi.org/10.1371/journal.pone.0243894.t006





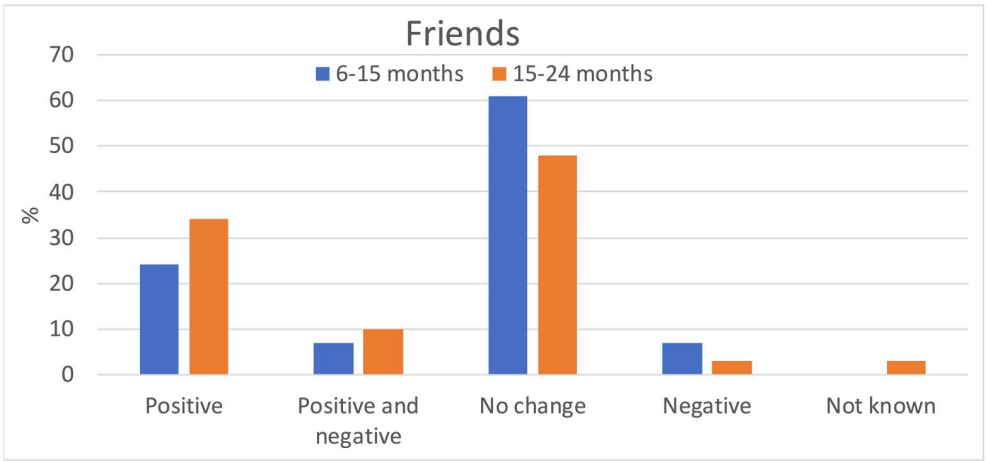


Fig 1. Ratings of change in life overall, mood and friendships at 6–15 months (n = 41) and 15–24 months (n = 29).

https://doi.org/10.1371/journal.pone.0243894.g001

experiencing more mood swings or feeling low. Findings at 15–24 months were similar. The most common negative change was reduced energy levels, reported by 29% at 6-15m and 38% at 15-24m.

Young people's reports of change in family and peer relationships were predominantly positive or neutral at both time points. Positive changes included feeling closer to the family,

Table 7. Adverse events reported across the study.

Participants	0-6m	7-12m	13-24m	25+m
	n = 44	n = 44	n = 36	n = 24
	n (%)	*n (%)*	*n (%)*	*n (%)*
Mild headaches or hot flushes	11 (25%)	10 (23%)	8 (22%)	4 (17%)
Moderate or severe headaches and hot flushes	2 (5%)	4 (9%)	1 (3%)	0
Fatigue—mild	2 (5%)	3 (7%)	3 (8%)	1 (4%)
Fatigue–moderate or severe	0	0	0	0
Mood swings	1 (2%)	0	0	0
Weight gain	1 (2%)	0	1 (3%)	0
Sleep problems	1 (2%)	0	1 (3%)	0
Other events	0	0	0	0
Total events recorded*	18	17	14	5

* individuals may have more than 1 event.

https://doi.org/10.1371/journal.pone.0243894.t007

feeling more accepted and having fewer arguments. Those reporting both positive and negative change reported feeling closer to some family members but not others. At 6–15 months, negative family changes were largely from family members not accepting their trans status or having more arguments. But by 15–24 months only one young person reported this. Improved relationships with peers related to feeling more sociable or confident and widening their circle of friends; negative changes related to bullying or disagreements at school. Again, at 15–24 months only one young person reported negative change, related to feelings of not trusting friends.

At 6–15 months, changes in gender role were reported by 66% as positive, including feeling more feminine/masculine, living in their preferred gender identity in more (or all) areas of life and feeling more secure in their gender identity, with no negative change reported. At 15–24 months, most reported no change although 41% reported positive changes including experimenting more with physical appearance and changing their details on legal documents.

All young people affirmed at each interview that they wished to continue with GnRHa treatment. Note that this was also the case when asked routinely at medical clinics (excepting those who briefly ceased GnRHa as noted above).

Adverse events. Adverse events are shown in Table 7. All adverse events were minor and anticipated, i.e. they were previously described in study participant information and/or noted in the triptorelin medication package inserts. Anticipated adverse events were common in the first two years, particularly mild headaches or hot flushes which were reported in 25% at 0-6m, 23% at 7-12m and 22% at 13-24m. Moderate or severe headaches and/or hot flushes were uncommon. Birth-registered females with distressing headaches or hot flushes were offered 'add-back' oestrogen therapy, and two accepted treatment briefly with very small doses of oestradiol, which was effective in reducing symptoms. Mild fatigue was reported by 5–8% over the first two years and no participants reported moderate or severe fatigue. Sleep problems, mood swings and weight gain were reported by very small numbers and in each case symptoms were mild. Adverse events were less common after 12 months of treatment.

Discussion

We report the short and medium-term outcomes of a prospective cohort of 44 young people with persistent and severe GD treated with GnRHa resulting in pubertal suppression from mid-puberty for 1–4 years. Young people were considered for recruitment after lengthy

assessment, spending an average of 2 years and up to 6 years within the GIDS psychological service before being referred to the endocrine clinic for assessment to enter the study. Medical assessment found no endocrine abnormalities at baseline. GnRHa treatment started in the majority of participants in later stages of puberty, with 57% in puberty stages 4 and 5 and 79% of birth-registered females being post-menarcheal. After starting GnRHa all quickly achieved and maintained suppression of pubertal hormones and none experienced pubertal progression. At the end of the study, 43 (98%) chose to start cross-sex hormones whilst one young person chose to stop GnRHa and continue with puberty consistent with their birth-registered sex.

As anticipated, pubertal suppression reduced growth that was dependent on puberty hormones, i.e. height and BMD. Height growth continued for those not yet at final height, but more slowly than for their peers so height z-score fell. Similarly for bone strength, BMD and BMC increased in the lumbar spine indicating greater bone strength, but more slowly than in peers so BMD z-score fell. These anticipated changes had been discussed with all participants before recruitment to the study. Young people experienced little change in mean weight or BMI z-score in the first two years. The rise in weight and BMI z-score at 36 months may represent a trend towards greater adiposity in those on GnRHa for a prolonged period, or reflect a higher baseline in this group.

Information on side-effects was available through routine reporting in medical clinics and in the participant experience interviews. Anticipated side effects of treatment were common, particularly mild symptoms directly related to suppression of sex hormones. Severe symptoms were uncommon. Fatigue or low energy was reported rarely in medical clinic assessments but frequently at interview (38% at 15-24m). The relationship of symptoms such as headaches, fatigue and sleep disturbance to GnRHa treatment is unclear as they are all very common in early adolescence [36,37], although a conservative perspective would regard them as side-effects of treatment.

Young people experienced little change in psychological functioning across the study. We found no differences between baseline and later outcomes for overall psychological distress as rated by parents and young people, nor for self-harm. Outcomes that were not formally tested also showed little change.

Participant experience of treatment as reported in interviews was positive for the majority, particularly relating to feeling happier, feeling more comfortable, better relationships with family and peers and positive changes in gender role. Smaller numbers reported having mixed positive and negative changes. A minority (12% at 6–15 months and 17% at 15–24 months) reported only negative changes, which were largely related to anticipated side effects. None wanted to stop treatment due to side effects or negative changes. We are not aware of comparative patient experience data from other cohorts.

The median age at consent in our study was very similar to that in the earliest published outcome study of mid-pubertal suppression using GnRHa treatment in Dutch young people (13.6 years) [24]. Similarly to this Dutch cohort, all but one of our participants elected to start cross-sex hormones after completing the GnRHa pathway. However they spent an average of 31 months on GnRHa compared with 23 months in the Dutch cohort [24]. In our study, the successful suppression of puberty and cessation of menses with GnRHa, the impact on height growth [4,16,38] and BMD [4,16] and the normality of liver and renal function through treatment were each consistent with previous reports [4,16].

Our findings that BMD increased over time in the lumbar spine but more slowly than in same age peers, resulting in a fall in z-score, are similar to others [4,14,39,40]. The fall in height-adjusted BMD z-score was consistent with but larger than the fall in height z-score. We found that birth-registered sex and pubertal status at baseline were not associated with later BMD. There is evidence that accretion of bone mass resumes and that BMD increases with the

start of cross-sex hormone therapy [4,14,39,41]. Future research needs to examine longer-term change in BMD in young people treated with mid-pubertal suppression.

We reported a range of adverse events previously described to be associated with pubertal suppression [42], with the exception of mild sleep disturbance although this is a known association with triptorelin use. As anticipated, the withdrawal of sex hormones produces symptoms such as headaches and lack of energy, although in the great majority (11 of 13 at 0–6 months; 10 of 14 at 7–12 months; 8 of 9 at 13–24 months) the symptoms were minor. Symptoms diminished over time as has previously been noted [4], and no young people chose to cease treatment due to the side-effects.

Our finding that 1 participant ceased pubertal suppression and did not commence cross-sex hormones is somewhat similar to the experience of one US cohort and a second Dutch cohort; Kuper et al. described that 2 of approximately 57 young people aged 10–15 years who commenced pubertal suppression treatment stopped this treatment without commencing cross-sex hormones [17]. Brik et al. reported that in a cohort of 137 young people who began GnRHa between 10 and 18 years and were followed until eligible to commence cross-sex hormones, 5 (3.6%) ceased treatment and did not later commence cross-sex hormones [19].

Three longitudinal studies from the Netherlands and the USA have examined psychological function over time in cohorts of young people treated with GnRHa and then cross-sex hormones [17,18,24], although the two US cohorts were of limited size. Our study adopted the same psychological outcome measures as the Dutch cohort, to facilitate comparison [24]. Mean baseline YSR scores in our cohort were similar to those previously reported in 141 young people aged 12–18 years from the London GIDS [43], and baseline CBCL and YSR scores were close to those at baseline from the original Dutch cohort [24]. A number of other studies have shown that young people with GD have higher scores on the CBCL or YSR than same-age population peers, and that they are similar to young people referred to clinical services for a range of mental health problems [44–46]. Population-based studies in America support higher baseline levels of mental health problems amongst young people with GD, with the prevalence of self-harm notably higher than for male or female peers [47,48]. Young people in our study had baseline YSR scores 0.7–1.0 SD higher than norms for age in comparable countries [29,46].

We found no evidence of change in psychological function with GnRHa treatment as indicated by parent report (CBCL) or self-report (YSR) of overall problems, internalising or externalising problems or self-harm. This is in contrast to the Dutch study which reported improved psychological function across total problems, externalising and internalising scores for both CBCL and YSR and small improvements in CGAS [24]. It also contrasts with a previous study from the UK GIDS of change in psychological function with GnRHa treatment in 101 older adolescents with GD (beginning $>$ 15.5 years) which reported moderate improvements in CGAS score over 12 months of GnRHa treatment [49]. CGAS scores in this previous study increased from 61 to 67 with GnRHa treatment, similar to those (63 at baseline, 66 at 24 months) in our study. Follow-up of the Kuper et al. cohort found non-significant changes in depression and anxiety scores in those (n = 25) who had only pubertal suppression treatment, although improvements were seen in the whole sample combining these with those receiving cross-sex hormones [17]. A second US cohort reported that in 23 young people who had received pubertal suppression (using GnRHa or anti-androgens in birth-registered males and either GnRHa or medroxyprogesterone in birth-registered females), there was a reduction in depression scores in birth-registered males but not females.

A recent large US survey found that those who received pubertal suppression in early or mid adolescence had lower odds of lifetime suicidal ideation when studied in adulthood compared with those who did not, regardless of whether they later received cross-sex hormones

and after adjustment for a range of confounding factors [50]. This implies an enduring benefit of pubertal suppression on psychological function, however the cross-sectional design and retrospective exposure classification means the findings require replication. Data are also available from other conditions in which GnRHa is used to suppress puberty during adolescence. A trial of GnRHa suppression of puberty during early adolescence in young people born small-for-gestational-age (SGA) who were also treated with human growth hormone (GH) reported that those treated with GnRHa had similar cognitive and psychological function in adult life to those treated only with GH [51].

The differences between our findings and the previous GIDS study re change in psychological function may relate simply to sample size. But why our findings differ from those of the Dutch study is unclear. They may relate to the timing of assessments; we assessed young people multiple times whereas in the Dutch study the second assessment was shortly before starting cross-sex hormone treatment. Alternatively, there may have been baseline differences in the two cohorts. Whilst some aspects of psychological function were similar, as noted above, the baseline CGAS scores were notably higher in the Dutch group (indicating better function). A previous international comparison study has found that young people aged 12–18 years with GD from the UK have higher scores indicating greater problems on the CBCL and YSR than those from the Netherlands, Belgium and Switzerland [52].

Psychological distress and self-harm are known to increase across early adolescence. Normative data show rising YSR total problems scores with age from age 11 to 16 years in non-clinical samples from a range of countries [29]. Self-harm rates in the general population in the UK and elsewhere increase markedly with age from early to mid-adolescence, being very low in 10 year olds and peaking around age 16–17 years [53–56]. Our finding that psychological function and self-harm did not change significantly during the study is consistent with two main alternative explanations. The first is that there was no change, and that GnRHa treatment brought no measurable benefit nor harm to psychological function in these young people with GD. This is consonant with the action of GnRHa, which only stops further pubertal development and does not change the body to be more congruent with a young person's gender identity. The second possibility is that the lack of change in an outcome that normally worsens in early adolescence may reflect a beneficial change in trajectory for that outcome, i.e. that GnRHa treatment reduced this normative worsening of problems. In the absence of a control group, we cannot distinguish between these possibilities. We aimed to use normative reference data to examine this issue. However age- and gender-standardised t-scores for ASEBA and other outcomes cannot answer this question as they cover a very broad age range (e.g. 12–18 years). We had anticipated that z-scores on the YSR available by calendar year for two comparable countries (Netherlands; Australia) might be informative however confidence intervals were too wide to draw reliable inferences.

Gender dysphoria and body image changed little across the study. This is consistent with some previous reports [24] and was anticipated, given that GnRHa does not change the body in the desired direction, but only temporarily prevents further masculinization or feminization. Other studies suggest that changes in body image or satisfaction in GD are largely confined to gender affirming treatments such as cross-sex hormones or surgery [57]. We found that birth-registered sex and baseline pubertal status were not associated with later psychological functioning on GnRHa, consistent with previous reports [24,49].

These data correct reports from a recent letter by Biggs [58] which used preliminary data from our study which were uncleaned and incomplete data used for internal reporting. In addition there were many statistical comparisons which inflated the risk of type 1 error. Our statistical analysis plan restricted testing all outcomes for differences by sex due to the type 1

error risk. Contrary to Biggs's letter, we found no evidence of reductions over time in any psychological outcomes, and no material differences by sex.

Strengths and limitations

Our study provides comprehensive data on this cohort during follow-up, with an anonymised dataset containing standardised scores deposited to allow other researchers to replicate our findings where data-sharing allows. The study size and uncontrolled design were key limitations. The small sample size limited our ability to identify small changes in outcomes. This was an uncontrolled observational study and thus cannot infer causality. Further, many of the outcomes studied here, including psychological function, self-harm and BMD, undergo normative changes by age and developmental stage during puberty that could confound any observed effect of GnRHa treatment in an uncontrolled study. The analysis plan aimed to take these issues into account as far as possible, however this particularly limits the potential for the study to show benefits or harms from treatment. However, some conclusions can be drawn. It is unlikely that the reported adverse events such as headaches do not relate directly to GnRHa treatment. Equally, given that there were no changes in psychological function and differences in point estimates were minimal for nearly all outcomes, it is unlikely that the treatment resulted in psychological harm. Observational studies are important sources of data on harms of treatment [59–61].

Our data are subject to a number of other limitations. This was an unfunded study undertaken within a clinical service and we were dependent on the clinical service for data collection. There were varying sample sizes for differing tests as some participants did not attend certain investigations and some follow-up medical tests were processed locally to patients; these data are reported as normal or otherwise. Missing items on psychological questionnaires resulted in some unusable data. Some young people found repeated completion of questionnaires about gender issues intrusive and refused to complete them at later follow-ups, as has been reported in other studies [62]. This questionnaire fatigue also affected parent responses. Scoring of psychological questionnaire data was rechecked at the completion of the study however this was not possible in very small numbers of participants in whom only scale scores rather than individual item data were preserved during data migration in hospital clinical information systems. In sensitivity analyses, repeat analysis of ASEBA psychological outcomes restricted to those with rescored data showed highly similar findings to the full sample (see S3 Table in S1 Appendix).

A more detailed qualitative evaluation of participant experience was not possible due to lack of interviewer time, and reporting of interview data was restricted to perceptions of positive or negative change and the giving of examples.

Implications and conclusions

Treatment of young people with persistent and severe GD aged 12–15 years with GnRHa was efficacious in suppressing pubertal progression. Anticipated effects of withdrawal of sex hormones on symptoms were common and there were no unexpected adverse events. BMD increased with treatment in the lumbar spine and was stable at the hip, and BMD z-score fell consistent with delay of puberty. Overall participant experience of changes on GnRHa treatment was positive. We identified no changes in psychological function, quality of life or degree of gender dysphoria.

The great majority of this cohort went on to start cross-sex hormones, as was hypothesized given the severity and continuation of their GD. However one young person did not, providing some evidence that development of gender identity continues on GnRHa treatment and

confirming the importance of continuing supportive psychological therapy to allow further exploration of gender identity and a range of future pathways whilst on GnRHa.

This cohort will be followed up longer term to examine physical and mental health outcomes into early adulthood. However larger and longer-term prospective studies using a range of designs are needed to more fully quantify the harms and benefits of pubertal suppression in GD and better understand factors influencing outcomes [3]. These are beginning to be funded in a number of countries [63].(https://logicstudy.uk) Given that pubertal suppression may be both a treatment in its own right and also an intermediate step in a longer treatment pathway, it is essential for such studies to examine benefits and harms across the longer pathway including pubertal suppression and initiation of cross-sex hormones.

Supporting information

S1 Appendix.
(DOCX)

S2 Appendix. Statistical analysis plan.
(DOCX)

Acknowledgments

We wish to thank the young people and families who participated in the study and the clinical teams at The Tavistock and Portman NHS Foundation Trust and UCL Hospitals NHS Foundation Trust.

We wish to acknowledge the inputs of Harriet Gunn, Claudia Zitz and Domenico di Ceglie for their work in formulating the study, collecting data and advising on the manuscript.

Author Contributions

Conceptualization: Polly Carmichael, Gary Butler, Elin M. Skageberg, Sophie Khadr, Russell M. Viner.

Data curation: Una Masic, Russell M. Viner.

Formal analysis: Tim J. Cole, Bianca L. De Stavola, Russell M. Viner.

Investigation: Gary Butler, Una Masic.

Methodology: Polly Carmichael, Bianca L. De Stavola, Elin M. Skageberg, Russell M. Viner.

Writing – original draft: Polly Carmichael, Gary Butler, Russell M. Viner.

Writing – review & editing: Polly Carmichael, Gary Butler, Una Masic, Tim J. Cole, Sarah Davidson, Sophie Khadr, Russell M. Viner.

References

1. Diagnostic and statistical manual of mental disorders (DSM-V). 5th ed. Arlington, VA: American Psychiatric Association; 2013.
2. International Statistical Classification of Diseases and Related Health Problems (ICD-11). Geneva: World Health Organisation, 2019.
3. Mahfouda S, Moore JK, Siafarikas A, Zepf FD, Lin A. Puberty suppression in transgender children and adolescents. Lancet Diabetes Endocrinol. 2017; 5(10):816–26. Epub 2017/05/27. https://doi.org/10.1016/S2213-8587(17)30099-2 PMID: 28546095.
4. Chew D, Anderson J, Williams K, May T, Pang K. Hormonal Treatment in Young People With Gender Dysphoria: A Systematic Review. Pediatrics. 2018; 141(4). Epub 2018/03/09. https://doi.org/10.1542/peds.2017-3742 PMID: 29514975.

5. Butler G, De Graaf N, Wren B, Carmichael P. Assessment and support of children and adolescents with gender dysphoria. Arch Dis Child. 2018; 103(7):631–6. Epub 2018/04/14. https://doi.org/10.1136/archdischild-2018-314992 PMID: 29650510.

6. Turban JL, Ehrensaft D. Research Review: Gender identity in youth: treatment paradigms and controversies. J Child Psychol Psychiatry. 2018; 59(12):1228–43. Epub 2017/10/27. https://doi.org/10.1111/jcpp.12833 PMID: 29071722.

7. Steensma TD, McGuire JK, Kreukels BP, Beekman AJ, Cohen-Kettenis PT. Factors associated with desistence and persistence of childhood gender dysphoria: a quantitative follow-up study. J Am Acad Child Adolesc Psychiatry. 2013; 52(6):582–90. Epub 2013/05/25. https://doi.org/10.1016/j.jaac.2013.03.016 PMID: 23702447.

8. Hembree WC, Cohen-Kettenis PT, Gooren L, Hannema SE, Meyer WJ, Murad MH, et al. Endocrine Treatment of Gender-Dysphoric/Gender-Incongruent Persons: An Endocrine Society Clinical Practice Guideline. J Clin Endocrinol Metab. 2017; 102(11):3869–903. Epub 2017/09/26. https://doi.org/10.1210/jc.2017-01658 PMID: 28945902.

9. Oliphant J, Veale J, Macdonald J, Carroll R, Johnson R, Harte M, et al. Guidelines for Gender Affirming Healthcare for Gender Diverse and Transgender Children, Young People and Adults in Aotearoa, New Zealand. N Z Med J. 2018; 131(1487):86–96. Epub 2018/12/14. PMID: 30543615.

10. Telfer MM, Tollit MA, Pace CC, Pang KC. Australian standards of care and treatment guidelines for transgender and gender diverse children and adolescents. Med J Aust. 2018; 209(3):132–6. Epub 2018/06/16. https://doi.org/10.5694/mja17.01044 PMID: 29902964.

11. Rafferty J. Ensuring Comprehensive Care and Support for Transgender and Gender-Diverse Children and Adolescents. Pediatrics. 2018; 142(4). Epub 2018/09/19. https://doi.org/10.1542/peds.2018-2162 PMID: 30224363.

12. Skordis N, Butler G, de Vries MC, Main K, Hannema SE. ESPE and PES International Survey of Centers and Clinicians Delivering Specialist Care for Children and Adolescents with Gender Dysphoria. Hormone research in paediatrics. 2018; 90(5):326–31. Epub 2019/01/30. https://doi.org/10.1159/000496115 PMID: 30695784.

13. Agana MG, Greydanus DE, Indyk JA, Calles JL Jr., Kushner J, Leibowitz S, et al. Caring for the transgender adolescent and young adult: Current concepts of an evolving process in the 21st century. Dis Mon. 2019; 65(9):303–56. Epub 2019/08/14. https://doi.org/10.1016/j.disamonth.2019.07.004 PMID: 31405516.

14. Klink D, Caris M, Heijboer A, van Trotsenburg M, Rotteveel J. Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. J Clin Endocrinol Metab. 2015; 100(2):E270–5. Epub 2014/11/27. https://doi.org/10.1210/jc.2014-2439 PMID: 25427144.

15. Cohen-Kettenis PT, Schagen SE, Steensma TD, de Vries AL, Delemarre-van de Waal HA. Puberty suppression in a gender-dysphoric adolescent: a 22-year follow-up. Arch Sex Behav. 2011; 40(4):843–7. Epub 2011/04/20. https://doi.org/10.1007/s10508-011-9758-9 PMID: 21503817.

16. Schagen SE, Cohen-Kettenis PT, Delemarre-van de Waal HA, Hannema SE. Efficacy and Safety of Gonadotropin-Releasing Hormone Agonist Treatment to Suppress Puberty in Gender Dysphoric Adolescents. J Sex Med. 2016; 13(7):1125–32. Epub 2016/06/19. https://doi.org/10.1016/j.jsxm.2016.05.004 PMID: 27318023.

17. Kuper LE, Stewart S, Preston S, Lau M, Lopez X. Body Dissatisfaction and Mental Health Outcomes of Youth on Gender-Affirming Hormone Therapy. Pediatrics. 2020; 145(4). Epub 2020/03/30. https://doi.org/10.1542/peds.2019-3006 PMID: 32220906.

18. Achille C, Taggart T, Eaton NR, Osipoff J, Tafuri K, Lane A, et al. Longitudinal impact of gender-affirming endocrine intervention on the mental health and well-being of transgender youths: preliminary results. International Journal of Pediatric Endocrinology. 2020; 2020(1):8. https://doi.org/10.1186/s13633-020-00078-2 PMID: 32368216

19. Brik T, Vrouenraets L, de Vries MC, Hannema SE. Trajectories of Adolescents Treated with Gonadotropin-Releasing Hormone Analogues for Gender Dysphoria. Arch Sex Behav. 2020; 49(7):2611–8. Epub 2020/03/11. https://doi.org/10.1007/s10508-020-01660-8 PMID: 32152785.

20. Hembree WC, Cohen-Kettenis P, Delemarre-van de Waal HA, Gooren LJ, Meyer WJ 3rd, Spack NP, et al. Endocrine treatment of transsexual persons: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab. 2009; 94(9):3132–54. Epub 2009/06/11. https://doi.org/10.1210/jc.2009-0345 PMID: 19509099.

21. Spack NP, Edwards-Leeper L, Feldman HA, Leibowitz S, Mandel F, Diamond DA, et al. Children and adolescents with gender identity disorder referred to a pediatric medical center. Pediatrics. 2012; 129(3):418–25. Epub 2012/02/22. https://doi.org/10.1542/peds.2011-0907 PMID: 22351896.

22. Meyer WJ 3rd. Gender identity disorder: an emerging problem for pediatricians. Pediatrics. 2012; 129 (3):571–3. Epub 2012/02/22. https://doi.org/10.1542/peds.2011-3696 PMID: 22351880.

23. Hewitt JK, Paul C, Kasiannan P, Grover SR, Newman LK, Warne GL. Hormone treatment of gender identity disorder in a cohort of children and adolescents. Med J Aust. 2012; 196(9):578–81. Epub 2012/05/25. https://doi.org/10.5694/mja12.10222 PMID: 22621149.

24. de Vries AL, Steensma TD, Doreleijers TA, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. J Sex Med. 2011; 8(8):2276–83. Epub 2010/07/22. https://doi.org/10.1111/j.1743-6109.2010.01943.x PMID: 20646177.

25. Cole TJ, Freeman JV, Preece MA. Body mass index reference curves for the UK, 1990. Arch Dis Child. 1995; 73(1):25–9. https://doi.org/10.1136/adc.73.1.25 PMID: 7639544

26. Kalkwarf HJ, Zemel BS, Gilsanz V, Lappe JM, Horlick M, Oberfield S, et al. The bone mineral density in childhood study: bone mineral content and density according to age, sex, and race. J Clin Endocrinol Metab. 2007; 92(6):2087–99. Epub 2007/02/22. https://doi.org/10.1210/jc.2006-2553 PMID: 17311856.

27. Zemel BS, Leonard MB, Kelly A, Lappe JM, Gilsanz V, Oberfield S, et al. Height adjustment in assessing dual energy x-ray absorptiometry measurements of bone mass and density in children. J Clin Endocrinol Metab. 2010; 95(3):1265–73. Epub 2010/01/28. https://doi.org/10.1210/jc.2009-2057 PMID: 20103654.

28. Crijnen AA, Achenbach TM, Verhulst FC. Comparisons of problems reported by parents of children in 12 cultures: total problems, externalizing, and internalizing. J Am Acad Child Adolesc Psychiatry. 1997; 36(9):1269–77. Epub 1997/09/18. https://doi.org/10.1097/00004583-199709000-00020 PMID: 9291729.

29. Verhulst FC, Achenbach TM, van der Ende J, Erol N, Lambert MC, Leung PW, et al. Comparisons of problems reported by youths from seven countries. Am J Psychiatry. 2003; 160(8):1479–85. Epub 2003/08/06. https://doi.org/10.1176/appi.ajp.160.8.1479 PMID: 12900311.

30. Aitken M, VanderLaan DP, Wasserman L, Stojanovski S, Zucker KJ. Self-Harm and Suicidality in Children Referred for Gender Dysphoria. J Am Acad Child Adolesc Psychiatry. 2016; 55(6):513–20. Epub 2016/05/31. https://doi.org/10.1016/j.jaac.2016.04.001 PMID: 27238070.

31. Van Meter AR, Algorta GP, Youngstrom EA, Lechtman Y, Youngstrom JK, Feeny NC, et al. Assessing for suicidal behavior in youth using the Achenbach System of Empirically Based Assessment. Eur Child Adolesc Psychiatry. 2018; 27(2):159–69. Epub 2017/07/28. https://doi.org/10.1007/s00787-017-1030-y PMID: 28748484.

32. Deutz MH, Geeraerts SB, van Baar AL, Dekovic M, Prinzie P. The Dysregulation Profile in middle childhood and adolescence across reporters: factor structure, measurement invariance, and links with self-harm and suicidal ideation. Eur Child Adolesc Psychiatry. 2016; 25(4):431–42. Epub 2015/08/01. https://doi.org/10.1007/s00787-015-0745-x PMID: 26226917.

33. Ravens-Sieberer U, Gosch A, Rajmil L, Erhart M, Bruil J, Power M, et al. The KIDSCREEN-52 quality of life measure for children and adolescents: psychometric results from a cross-cultural survey in 13 European countries. Value Health. 2008; 11(4):645–58. Epub 2008/01/09. https://doi.org/10.1111/j.1524-4733.2007.00291.x PMID: 18179669.

34. Lindgren TW, Pauly IB. A body image scale for evaluating transsexuals. Arch Sex Behav. 1975; 4 (6):639–56. Epub 1975/11/01. https://doi.org/10.1007/BF01544272 PMID: 1212093.

35. Shaffer D, Gould MS, Brasic J, Ambrosini P, Fisher P, Bird H, et al. A children's global assessment scale (CGAS). Arch Gen Psychiatry. 1983; 40(11):1228–31. Epub 1983/11/01. https://doi.org/10.1001/archpsyc.1983.01790100074010 PMID: 6639293.

36. Viner RM, Clark C, Taylor S, Bhui K, Klineberg E, Head J, et al. Risk factors for persistent fatigue in adolescents: A population-based study. Journal of Adolescent Health. 2006; 38(2):113–4. https://doi.org/10.1016/j.jadohealth.2005.11.080

37. Krogh AB, Larsson B, Linde M. Prevalence and disability of headache among Norwegian adolescents: A cross-sectional school-based study. Cephalalgia. 2015; 35(13):1181–91. Epub 2015/02/28. https://doi.org/10.1177/0333102415573512 PMID: 25720767.

38. Ghelani R, Lim C, Brain C, Fewtrell M, Butler G. Sudden sex hormone withdrawal and the effects on body composition in late pubertal adolescents with gender dysphoria. J Pediatr Endocrinol Metab. 2019. Epub 2019/12/14. https://doi.org/10.1515/jpem-2019-0045 PMID: 31834861.

39. Vlot MC, Klink DT, den Heijer M, Blankenstein MA, Rotteveel J, Heijboer AC. Effect of pubertal suppression and cross-sex hormone therapy on bone turnover markers and bone mineral apparent density (BMAD) in transgender adolescents. Bone. 2017; 95:11–9. Epub 2016/11/16. https://doi.org/10.1016/j.bone.2016.11.008 PMID: 27845262.

40. Joseph T, Ting J, Butler G. The effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria: findings from a large national cohort. J Pediatr Endocrinol Metab. 2019; 32(10):1077–81. Epub 2019/09/01. https://doi.org/10.1515/jpem-2019-0046 PMID: 31472062.

41. Rothman MS, Iwamoto SJ. Bone Health in the Transgender Population. Clin Rev Bone Miner Metab. 2019; 17(2):77–85. Epub 2019/08/28. https://doi.org/10.1007/s12018-019-09261-3 PMID: 31452648.

42. Panagiotakopoulos L. Transgender medicine—puberty suppression. Rev Endocr Metab Disord. 2018; 19(3):221–5. Epub 2018/08/17. https://doi.org/10.1007/s11154-018-9457-0 PMID: 30112593.

43. Skagerberg E, Davidson S, Carmichael P. Internalizing and Externalizing Behaviors in a Group of Young People with Gender Dysphoria. Int J Transgenderism. 2013; 13(3):105–12. https://doi.org/10.1080/15532739.2013.822340

44. Zucker KJ. Adolescents with Gender Dysphoria: Reflections on Some Contemporary Clinical and Research Issues. Arch Sex Behav. 2019; 48(7):1983–92. Epub 2019/07/20. https://doi.org/10.1007/s10508-019-01518-8 PMID: 31321594.

45. Zucker KJ, Bradley SJ, Owen-Anderson A, Kibblewhite SJ, Wood H, Singh D, et al. Demographics, behavior problems, and psychosexual characteristics of adolescents with gender identity disorder or transvestic fetishism. J Sex Marital Ther. 2012; 38(2):151–89. Epub 2012/03/07. https://doi.org/10.1080/0092623X.2011.611219 PMID: 22390530.

46. Levitan N, Barkmann C, Richter-Appelt H, Schulte-Markwort M, Becker-Hebly I. Risk factors for psychological functioning in German adolescents with gender dysphoria: poor peer relations and general family functioning. Eur Child Adolesc Psychiatry. 2019; 28(11):1487–98. Epub 2019/03/17. https://doi.org/10.1007/s00787-019-01308-6 PMID: 30877477.

47. Becerra-Culqui TA, Liu Y, Nash R, Cromwell L, Flanders WD, Getahun D, et al. Mental Health of Transgender and Gender Nonconforming Youth Compared With Their Peers. Pediatrics. 2018; 141(5). Epub 2018/04/18. https://doi.org/10.1542/peds.2017-3845 PMID: 29661941.

48. Taliaferro LA, McMorris BJ, Rider GN, Eisenberg ME. Risk and Protective Factors for Self-Harm in a Population-Based Sample of Transgender Youth. Arch Suicide Res. 2019; 23(2):203–21. Epub 2018/02/21. https://doi.org/10.1080/13811118.2018.1430639 PMID: 29461934.

49. Costa R, Dunsford M, Skagerberg E, Holt V, Carmichael P, Colizzi M. Psychological Support, Puberty Suppression, and Psychosocial Functioning in Adolescents with Gender Dysphoria. J Sex Med. 2015; 12(11):2206–14. Epub 2015/11/12. https://doi.org/10.1111/jsm.13034 PMID: 26556015.

50. Turban JL, King D, Carswell JM, Keuroghlian AS. Pubertal Suppression for Transgender Youth and Risk of Suicidal Ideation. Pediatrics. 2020. Epub 2020/01/25. https://doi.org/10.1542/peds.2019-1725 PMID: 31974216.

51. Goedegebuure WJ, van der Steen M, de With JL, Hokken-Koelega A. Cognition, Health-Related Quality of Life, and Psychosocial Functioning After GH/GnRHa Treatment in Young Adults Born SGA. J Clin Endocrinol Metab. 2018; 103(11):3931–8. Epub 2018/08/24. https://doi.org/10.1210/jc.2018-01463 PMID: 30137415.

52. de Graaf NM, Cohen-Kettenis PT, Carmichael P, de Vries ALC, Dhondt K, Laridaen J, et al. Psychological functioning in adolescents referred to specialist gender identity clinics across Europe: a clinical comparison study between four clinics. Eur Child Adolesc Psychiatry. 2018; 27(7):909–19. Epub 2017/12/20. https://doi.org/10.1007/s00787-017-1098-4 PMID: 29256158.

53. Nock MK, Green JG, Hwang I, McLaughlin KA, Sampson NA, Zaslavsky AM, et al. Prevalence, correlates, and treatment of lifetime suicidal behavior among adolescents: results from the National Comorbidity Survey Replication Adolescent Supplement. JAMA Psychiatry. 2013; 70(3):300–10. Epub 2013/01/11. https://doi.org/10.1001/2013.jamapsychiatry.55 PMID: 23303463.

54. Jung KY, Kim T, Hwang SY, Lee TR, Yoon H, Shin TG, et al. Deliberate Self-harm among Young People Begins to Increase at the Very Early Age: a Nationwide Study. Journal of Korean medical science. 2018; 33(30):e191. Epub 2018/07/24. https://doi.org/10.3346/jkms.2018.33.e191 PMID: 30034304.

55. Morey Y, Mellon D, Dailami N, Verne J, Tapp A. Adolescent self-harm in the community: an update on prevalence using a self-report survey of adolescents aged 13–18 in England. J Public Health (Oxf). 2017; 39(1):58–64. Epub 2016/02/20. https://doi.org/10.1093/pubmed/fdw010 PMID: 26892623.

56. Stallard P, Spears M, Montgomery AA, Phillips R, Sayal K. Self-harm in young adolescents (12–16 years): onset and short-term continuation in a community sample. BMC psychiatry. 2013; 13:328. Epub 2013/12/04. https://doi.org/10.1186/1471-244X-13-328 PMID: 24294921.

57. Becker I, Auer M, Barkmann C, Fuss J, Moller B, Nieder TO, et al. A Cross-Sectional Multicenter Study of Multidimensional Body Image in Adolescents and Adults with Gender Dysphoria Before and After Transition-Related Medical Interventions. Arch Sex Behav. 2018; 47(8):2335–47. Epub 2018/08/09. https://doi.org/10.1007/s10508-018-1278-4 PMID: 30088234.

58. Biggs M. Gender Dysphoria and Psychological Functioning in Adolescents Treated with GnRHa: Comparing Dutch and English Prospective Studies. Arch Sex Behav. 2020; 49(7):2231–6. Epub 2020/07/01. https://doi.org/10.1007/s10508-020-01764-1 PMID: 32594279.

59. Vandenbroucke JP. What is the best evidence for determining harms of medical treatment? CMAJ: Canadian Medical Association journal = journal de l'Association medicale canadienne. 2006; 174 (5):645–6. Epub 2006/03/01. https://doi.org/10.1503/cmaj.051484 PMID: 16505461.

60. Anglemyer A, Horvath HT, Bero L. Healthcare outcomes assessed with observational study designs compared with those assessed in randomized trials. The Cochrane database of systematic reviews. 2014;(4):MR000034. Epub 2014/05/02. https://doi.org/10.1002/14651858.MR000034.pub2 PMID: 24782322.

61. Golder S, Loke YK, Bland M. Meta-analyses of adverse effects data derived from randomised controlled trials as compared to observational studies: methodological overview. PLoS Med. 2011; 8(5): e1001026. Epub 2011/05/12. https://doi.org/10.1371/journal.pmed.1001026 PMID: 21559325.

62. Olson-Kennedy J, Chan YM, Rosenthal S, Hidalgo MA, Chen D, Clark L, et al. Creating the Trans Youth Research Network: A Collaborative Research Endeavor. Transgend Health. 2019; 4(1):304–12. Epub 2019/11/09. https://doi.org/10.1089/trgh.2019.0024 PMID: 31701011.

63. Olson J, Chan Y-M, Garofalo R, Spack NP, Chen D, Clark L, et al. Impact of Early Medical Treatment for Transgender Youth: Protocol for the Longitudinal, Observational Trans Youth Care Study. JMIR research protocols. 2019; 8(7):e14434. https://doi.org/10.2196/14434 PMID: 31290407

Nordic Journal of Psychiatry

ISSN: 0803-9488 (Print) 1502-4725 (Online) Journal homepage: https://www.tandfonline.com/loi/ipsc20

Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria

Riittakerttu Kaltiala, Elias Heino, Marja Työläjärvi & Laura Suomalainen

To cite this article: Riittakerttu Kaltiala, Elias Heino, Marja Työläjärvi & Laura Suomalainen (2020) Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria, Nordic Journal of Psychiatry, 74:3, 213-219, DOI: 10.1080/08039488.2019.1691260

To link to this article: https://doi.org/10.1080/08039488.2019.1691260

Published online: 25 Nov 2019.

NORDIC JOURNAL OF PSYCHIATRY
2020, VOL. 74, NO. 3, 213–219
https://doi.org/10.1080/08039488.2019.1691260

Taylor & Francis
Taylor & Francis Group

ARTICLE

Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria

Riittakerttu Kaltiala[a,b,c], Elias Heino[b], Marja Työläjärvi[a] and Laura Suomalainen[d]

[a]Department of Adolescent Psychiatry, Tampere University Hospital, Tampere, Finland; [b]Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland; [c]Vanha Vaasa Hospital, Vaasa, Finland; [d]Department of Adolescent Psychiatry, Helsinki University Hospital, Helsinki, Finland

ABSTRACT

Purpose: To assess how adolescent development progresses and psychiatric symptoms develop among transsexual adolescents after starting cross-sex hormone treatment.

Materials and methods: Retrospective chart review among 52 adolescents who came into gender identity assessment before age 18, were diagnosed with transsexualism and started hormonal gender reassignment. The subjects were followed over the so-called real-life phase of gender reassignment.

Results: Those who did well in terms of psychiatric symptoms and functioning before cross-sex hormones mainly did well during real-life. Those who had psychiatric treatment needs or problems in school, peer relationships and managing everyday matters outside of home continued to have problems during real-life.

Conclusion: Medical gender reassignment is not enough to improve functioning and relieve psychiatric comorbidities among adolescents with gender dysphoria. Appropriate interventions are warranted for psychiatric comorbidities and problems in adolescent development.

ARTICLE HISTORY
Received 3 June 2019
Revised 24 October 2019
Accepted 6 November 2019

KEYWORDS
Gender dysphoria; transsexualism; adolescence; adolescent development; cross-sex hormones

Introduction

Adolescence starts from puberty and ends approximately ten years later with the consolidation of adulthood personality structures [1,2]. The upsurge of steroid hormones in puberty initiates the maturation of the reproductive system and secondary sexual characteristics, and also vast structural and functional developments in the brain [3]. These biological changes are accompanied by extensive cognitive, emotional and social changes characteristic of adolescent development. The psychosocial developmental tasks of adolescence comprise sexual maturation (including adopting to the sexually maturing body and becoming capable of mutually satisfying, reciprocal romantic and sexual relationships), achieving independence from parents, and assuming an identity and responsible social role [1,4–6].

Gender Dysphoria (GD) refers to a marked discrepancy between the experienced gender and biological sex, causing clinically significant distress or impairment in functioning (DSM-5) [7]. Individuals with GD often wish to obtain hormonal and surgical treatments to align their body with the experiences gender. In ICD-10 the corresponding diagnosis is Transsexualism (ICD-10) [8].

Favourably progressing adolescent development manifests in the adolescent's functioning in relation to her/his own sexually maturing body, parents, peers, romance and sexuality, and school/future career [4,9,10]. The literature exploring adolescent development and functioning among adolescents with gender dysphoria and/or transgender identity is scarce and scattered. The sexually maturing body is a core challenge for adolescents suffering from gender dysphoria. A recent review suggested that adolescent gender dysphoria/transgender identity is associated with both negative (rejection, bullying) and positive (closer relationship, inclusion, attention) features in parent and peer relationships, both delayed and advanced for age or risky sexual behaviours, and with school-related challenges that are primarily assumed to relate to prejudice and peer rejection [10].

Psychiatric comorbidities, particularly depression, anxiety disorders and autism spectrum disorders as well as suicidality and self-harming behaviours are common among adolescents seeking gender reassignment [10]. Psychiatric comorbidities cannot automatically be assumed to be secondary to gender dysphoria [11] and do not necessarily remit due to sex reassignment [12].

During the past ten years the number of adolescents contacting gender identity services in order to seek for medical gender reassignment has increased across Western countries [13–16]. The reasons for this are not known [10].

Medical approaches to adolescent gender dysphoria may comprise halting/delaying the physical maturation (puberty blocking), and cross-sex hormonal treatments. Surgical treatments are mainly available for legal adults [17,18]. Medical gender reassignment is expected to alleviate gender dysphoria, psychiatric comorbidities and related

CONTACT Riittakerttu Kaltiala riittakerttu.kaltiala@tuni.fi Department of Adolescent Psychiatry, Tampere University Hospital, Box 2000, Tampere, 33521, Finland

psychosocial problems. Initial studies have suggested that puberty blocking with GnRH analogues may reduce psychiatric symptoms and improve functioning in gender dysphoric adolescents [19,20], but follow-up studies assessing the effectiveness and safety of hormonal interventions initiated during the developmental years are, however, scarce and biased by methodological problems to the extent that a recent meta-analysis concluded that they must be considered experimental [21,22]. There is an urgent need for follow-up studies on the outcomes of gender identity based hormonal interventions initiated during adolescent development.

The aim of this study was to evaluate the adolescent development of young people diagnosed with transsexualism and offered cross-sex hormonal interventions in one of the two gender identity units for minors in the period 2011–2017. We set out to evaluate the psychosocial functioning and need for psychiatric treatment of this patient group during the gender identity diagnostic phase and after about a year on cross-sex hormone treatment. We expected to see improvements in psychosocial functioning and a decrease in need for psychiatric treatment after starting the hormonal treatment that results in the desired changes in secondary sexual characteristics, which expectedly alleviates gender dysphoria.

Materials and methods

In Finland, the gender identity assessments required in order to proceed to medical sex reassignment interventions are centralized to two of the five university hospitals in the country. After the diagnostic assessments, legal sex change can take place after a period of about a year on cross-sex hormonal treatments, the so-called real-life phase of living in the desired role. Diagnostic assessments in Finnish health care take place according to ICD-10 (8). Legal sex change and surgical treatments require the patient to have achieved legal majority (18 years). To proceed to legal sex change, the patient has to obtain a certificate from the gender identity unit that carried out the primary diagnostic assessments and from the other gender identity service (second opinion). Gender identity assessments for minors were initiated in 2011.

The study comprises a retrospective chart review of adolescents referred to one of the two gender identity service facilities for minors in Finland (Tampere University Hospital, Department of Adolescent Psychiatry) before age 18, who had been diagnosed with transsexualism and proceeded to cross-sex hormonal treatments and who had completed a follow-up of approximately a year after starting on cross-sex hormones (real-life phase).

The assessments conducted by the gender identity team comprise structured and free format assessments and interviews by a multi-disciplinary team and an evaluation of the adolescent's existing psychiatric and medical files [11]. Two of the authors (RK, MT) were involved in the clinical assessments of all the gender-referred adolescents during the study period. The research data was collected retrospectively from the case files by a junior researcher (EH) trained and supervised by the first author. All information available after the clinical evaluations was used and the data was collected with help of a structured data collection form until the referral for the second opinion in the other adolescent gender identity unit was written. The study received approval from the ethics committee of Tampere University Hospital.

Between 2011 and 2017, 57 adolescents had been diagnosed with F64.0, transsexualism, and had been offered an opportunity to start hormonal sex reassignment. One of them did not want any treatment, two withdrew and two had started hormonal treatments but had not yet completed the real-life phase at the end of 2017. Thus, 52 patients were included in the study. Of these 11 were birth assigned males (transfemales) and 41 birth assigned females (transmales). They had a mean (sd) age of 18.1 (1.1) years at diagnosis, range 15.2–19.9 years (no difference between sexes).

Measures

Indicators of adolescent development

Adolescent development was evaluated in terms of age-appropriate living arrangements, peer relationships, school/work participation, romantic involvement, competence in managing everyday matters and need for psychiatric treatment.

Living arrangements were classified as (1) living with at least one parent/guardian, (2) living in a boarding school, with an adult relative, in some form of supported accommodation or the like, where supervision and guidance by a responsible adult is provided, (3) independently alone or in a shared household with a peer, (4) with a romantic partner. In the analyses dichotomized living arrangements (a) during gender identity assessment and (b) during the real-life phase living with (a) parent(s)/guardian(s) vs. in other arrangements. In Finnish culture, minors younger than 18 years usually live in the parental home, but leaving the parental home takes place earlier than in the majority of EU countries. Of young people aged 20–24, about a fourth are living in the parental home in Finland [23,24]

Peer relationships were classified as follows: (1) socializes with friends in leisure time, outside of activities supervised by adults, (2) socializes with peers only at school or in the context of rehabilitative activity, (3) spends time close to peers, for example in school or rehabilitative activity, but does not connect with them, (4) does not meet peers at all. In the analyses, peer relationships during (a) gender identity assessment and (b) the real-life phase were dichotomized to age-appropriate (normative) [1] vs. restricted or lacking [2–4].

School/work participation was classified as (1) age appropriate participation in mainstream curriculum, progresses without difficulties, (2) participates in mainstream curriculum with difficulty, (3) participates in rehabilitative educational or work activity, (4) not involved in education and working life. Age-appropriate participation during [1] was recorded if the adolescent attended mainstream secondary education or upper secondary education at a regular rate (a class per year in comprehensive school; has not changed more than once

between tracks in upper secondary education) or had proceeded to work life after completing vocational education. Participation with difficulty [2] was recorded if the adolescent was enrolled in mainstream education but had to repeat a class, studied with special arrangements (for example, in a special small group), or followed some form of adjusted curriculum. In the analyses, school/work life during (a) gender identity assessment and (b) real-life phase was dichotomized to normative [1] vs. any other (2, 3 or 4).

Romantic involvement was recorded (1) has or has had a dating or steady relationship, not only online, (2) has had a romantic relationship only online, (3) has not had dating or steady relationships. In the analyses we compared has or has had [1] vs. has not had [2,3] a dating or steady relationship during (a) gender identity assessment and (b) real-life phase. Sexual history was recorded in more detail in case histories during gender identity assessment, and for this period we also collected the experiences of (French) kissing (yes/no), intercourse (yes/no) and experience of any genitally intimate contact with a partner (petting under clothes or naked, intercourse, oral sex) (yes/no).

In recording age-appropriate competence in managing everyday matters we expected that early adolescents (up to 14 years) would be able, for example, to do shopping and travel alone on local public transport, and to help with household duties assigned by their parents. Middle adolescents (15–17 years) were further assumed, for example, to be able make telephone calls in matters important to them (for example, when seeking a summer job), to deal with school-related issues with school personnel without parental participation, to select and start new hobbies independently and to fulfil their role in summer jobs and in similar responsibilities of young people. Late adolescents (18+ years), legally adults, were expected to have, in addition to the above, competence to talk to authorities such as professionals in health and social services, employment or educational institutions, to deal with banks or health insurance, to manage their financial issues and to manage their housekeeping if they chose to move to live independently of parents/guardians. Competence in managing everyday matters was recorded as follows: (1) the adolescent is able to cope age-appropriately outside home, (2) the adolescent needs support in age-appropriate matters outside home but functions age-appropriately in the home (manages her/his own hygiene, clothing and nutrition, participates in (younger subjects) or takes responsibility for (older subjects) housekeeping) and (3) the adolescent's functioning is inadequate both at home and outside home. In the analyses we focused in being age-appropriately able cope with matters outside of the home [1] vs. not [2,3].

Psychiatric disorders (depression, anxiety, suicidality/self-harm, conduct problems, substance abuse problems, psychoses, ADHD, autism, eating disorders) were recorded a) if they had required specialist level psychiatric treatment during or before the gender identity assessment, (i.e. the adolescent was in treatment, or treatment had been recommended but the adolescent refused it) and b) if they required specialist level psychiatric treatment during the real-life phase (i.e. the adolescent was in treatment or the psychiatrist in the gender identity unit recorded that treatment was recommended or made a referral to psychiatric treatment irrespective of whether or not the adolescent complied with the recommendation).

Statistical analyses

Distributions of variables illustrating adolescent development are given for (a) the time of the gender identity assessment and (b) the real-life phase. Differences in proportions displaying age-appropriate functioning were compared using chi-square statistics/Fisher's exact test as appropriate. Cross-tabulations with chi-square statistics/Fisher's exact test as appropriate were used to explore functioning on a domain during the real-life phase according to functioning therein during assessment (i.e. school/work during real-life phase according to school/work during assessment etc.).

Need for specialist level psychiatric treatment before or during the gender identity assessment and during the real-life phase was compared using cross-tabulations with chi square statistics. Similarly, need for treatment according to the nine disorder dimensions recorded was compared between the two time periods. The associations between need for specialist level psychiatric treatment a) before or during the gender identity assessment, and b) during the real life-phase and functioning in the domains studied were explored using cross-tabulation with chi-square statistics/Fisher's exact test where appropriate.

The role of sex/gender and age were analysed by logistic regression. Functioning in peer relationships, school/work, managing everyday matters and dating/going steady were entered each in turn as the dependent variable with age and sex/gender as independent variables. Odds Ratios (OR) with 95% confidence intervals (CI) were calculated.

Results

Adolescent development and need for treatment during assessment and during real-life phase

During the gender identity assessment, three quarters of the adolescents lived with their parents. About three out of five displayed age-appropriate progress in school/work, four out of five functioned age-appropriately in dealing with matters outside home, and almost all had normative peer contacts. About three out of five had experienced dating/steady relationships before the end of the gender identity assessment (Table 1). In more detail about sexual development, 83% (43/52) had been in love/had a crush on someone, 56% (29/52) had experienced kissing, 8% (4/52) intercourse and 64% (33/52) any genitally intimate sexual contact with a partner by the end of the gender identity assessment.

During and before the gender identity assessment, half of the adolescents required specialist level psychiatric treatment, most commonly because of depression, anxiety, and suicidality/self-harm (Table 2).

In the end of the real-life phase, a majority had moved on to live independently of parents/guardians. The shares of

Table 1. Functioning in different domains of adolescent development during gender identity assessment and real-life phase among 52 young people diagnosed with transsexualism after starting gender identity assessments before age 18 [% (n/N)].

	During gender identity assessment	During real life phase	*p* Value
Living with parent(s)/guardians	73% (38/52)	40% (21/50)	0.001
Normative peer contacts	89% (46/52)	81% (42/52)	<0.001
Progresses normatively in school/ work	64% (33/52)	60% (31/52)	0.69
Has had dating or steady relationships	62% (32/50)	58% (30/52)	0.51
Is age-appropriately able to dealt with matters outside of the home	81% (42/52)	81% (42/52)	1.0

Table 2. Need for specialist level psychiatric treatment, and disorder/symptom dimensions requiring this treatment during and before gender identity assessment, and during real life phase [% (n/N)].

	During and before gender identity assessment	During real life phase	*p* Value
Need for psychiatric treatment	50% (26/52)	46% (24/51)	0.77
Need for treatment due to …			
depression	54% (28/52)	15% (8/52)	<0.001
anxiety	48% (25/52)	15% (8/52	<0.001
suicidality/self-harm	35% (18/52)	4% (2/52)	<0.001
conduct problems/antisocial	14% (7/52)	6% (3/52)	0.18
psychotic symptoms/psychosis	2% (1/52)	4% (2/52)	0.56
substance abuse	4% (2/52)	2% (1/52)	0.56
autism	12% (6/52)	6% (3/52	0.30
ADHD	10% (5/52)	2% (1/52)	0.09
eating disorder	2% (1/52)	2% (1/52)	1.0

those progressing age-appropriately in school/work, dealing age-appropriately with matters outside of home and being involved in dating/steady relationships did not change from the assessment phase to the end of the real-life phase. The proportion of those functioning age-appropriately in peer relationships decreased from the assessment period to the real-life phase (Table 1). The share of those requiring specialist level psychiatric treatment during real-life due to any reason was similar to that during and before the assessment, but treatment needs due to depression, anxiety and suicidality/self-harm had diminished (Table 2).

Changes within different domains of functioning

Of those adolescents with-age appropriate peer contacts during assessment (46/52), 91% (42/46) continued to have age-appropriate peer contacts during the real-life phase while 9% (4/46) no longer had these. Of those with difficulties in peer contacts (6/52), all continued to have difficulties in this field. ($p < 0.001$)

Of those who progressed age-appropriately at school (working life) during assessment (33/52), 85% (28/33) continued to do so during the real-life phase, but 15% (5/33) did not. Of those with problems at school (work) (19/52), 84% (16/19) continued to have problems, but 16% (3/19) ceased to have problems in this field. ($p < 0.001$)

Of those who had had age-appropriate skills in dealing with matters outside home (42/52), 88% (37/42) continued to be able to do so but 12% (5/42) functioned below the age-appropriate level during the real-life phase. Of those who had had difficulties in dealing with matters outside home (10/52), half (5/10) continued to do so, but half (5/10) no longer had problems in this field ($p = 0.02$).

Of those who had experiences of dating/steady relationships during the assessment (32/50), 66% (21/32) had dating/steady relationships during the real-life phase, and 34% (11/32) did not. Of those who had not had any dating/steady relationships by the end of the gender identity assessment, 44% (8/18) had and 56% (10/18) did not have these during the real-life phase. ($p = 0.12$)

Of those not needing psychiatric treatment before or during the assessment (26/52), 73% (19/26) did not need any during the real-life phase but in 27% (7/26), a need had emerged. Of those who had needed (25/51) psychiatric treatment during or before the assessment, 68% (17/25) still needed it during the follow-up but 32% (8/25) did not. ($p = 0.004$)

The role of psychiatric comorbidities for functioning during real life

Need for psychiatric treatment before or during the real-life phase was not associated with functioning in peer relationships or romantic relationships during the real-life phase. Those needing psychiatric treatment before or during gender identity assessment were more likely to not function age-appropriately in school/work (47% (15/32) vs. 82% (14/17) functioned well, $p = 0.02$), and borderline significantly less likely to cope well with managing everyday matters outside home (72% (23/32) vs. 94% (16/17) managed well, $p = 0.06$) during the real-life phase.

Concurrent need for psychiatric treatment during the real-life phase was associated with a smaller proportion functioning well at school/work [42% (10/24) vs. 74% (20/27), $p = 0.02$] and in taking care of everyday matters [67% (16/24) vs. 93% (25/27), $p = 0.02$].

No associations were found between age and sex (gender) and functional outcomes.

Discussion

The aim of this study was to assess the adolescent development of those adolescents who were diagnosed with transsexualism and offered cross-sex hormonal interventions during the subsequent real-life phase, when the cross-sex hormonal treatment was initiated and started to produce the desired changes in physical appearance. Moving to live independently, relationships with peers, romantic involvement, ability to take care of everyday issues age-appropriately outside home and need for psychiatric treatment were assessed as proxies for adolescent development. Earlier empirical research on outcomes of medical sex reassignment interventions initiated during developmental years was scarce and offered little advice on the impact of treatments on adolescent development [10,21,22].

We observed that the majority of the adolescents diagnosed with transsexualism and offered cross- sex hormonal treatments displayed age-appropriate functioning in the

domains studied during the gender identity assessment, as is to be expected given that severe psychopathology and markedly lowered functioning may complicate the possibilities to assess identity achievement and may constitute a contraindication for medical treatment. Nevertheless, a considerable share also had difficulties in different domains of functioning. What is more, even if the majority also functioned well in the domains studied during the first year on cross-sex hormones, no statistically significant improvements in functioning were observed in the group as a whole, and in the domain of peer relationships the share of those with normative contacts decreased. This is in disagreement with earlier studies suggesting improved functioning and reduced psychiatric symptoms in adolescent onset hormonal treatment of gender dysphoria [19,20], and likely due to older age, more difficult psychopathology and different intervention (cross-sex hormones vs. GnRH analogues) in our sample. Our subjects were all post-pubertal and halting of development was thus not possible.

The majority of the adolescents diagnosed with transsexualism were still living in the parental home during the gender identity assessment, which is to be expected and culturally normal as they were in the age range of 15.2–19.9 years. During the subsequent real-life phase, the share of those living in the parental home decreased. This concurs with progression of adolescent development. Given the knowledge of normative timing of leaving the parental home in Finland [21,22], the increasing proportion of those no longer living with their parents likely indicates positive progress in adolescent development instead of, for example, negative parental reactions to sex reassignment, which has also been reported in the literature [25], particularly as most of those leaving the parental home went to live with romantic partners (data not shown). Due to excellent social security benefits, moving to live independently does not necessitate regular income from employment and is therefore not a proxy for good functioning in other domains of life.

The difficulties in peer relationships commonly reported among adolescents with transgender identities have been associated with prejudice and discrimination [10,26]. Anxiety disorders, particularly social anxiety, could relate both to victimization and distress created by not being able to satisfactorily present oneself according to one's perceived gender. With the appearance of the desired physical characteristics, passing in the desired role is expected to be facilitated and self-confidence to increase. Positive changes in connection with peers could be expected. However, of those who had difficulties in peer relationships during the gender identity assessment, all continued to have them in the follow-up, and almost one in ten of those functioning well in this domain during the assessment developed difficulties in follow-up. This was contrary to our expectations and suggests that difficulties in peer relationships cannot be attributed to difficulties in passing in the desired role.

About two out of five of the adolescents diagnosed with transsexualism had experienced dating or steady relationships by the end of the gender identity assessments, and an equal share during the real-life phase. For comparison, recent Finnish data on 15-year-old adolescents reveals that about a half of them have experienced dating/steady relationships (unpublished observation). Steady relationships in adolescence may be short and not dating/going steady exactly during the real-life phase cannot be taken as an indicator of delayed development. Earlier studies have shown that clinically referred adolescents with gender dysphoria display normative emotional development in regard to romance and dating but show slight delays in behavioural level sexual development [27,28]. Compared to earlier findings on all gender-referred adolescents, the adolescents now studied had experienced falling in love and dating/steady relationships about equally frequently but had slightly less often engaged in sexually intimate behaviours than both all gender referred adolescents and same aged adolescents in general population (Kaltiala-Heino et al. [28]). These observations do not suggest remarkable delays in sexual development. During the real-life phase, a considerable share also gained their first experiences of dating/steady relationships, which suggests favourable progression of adolescent development.

If the adolescents diagnosed with transsexualism had had difficulties at school/work as during the gender identity assessment, they mainly continued to have difficulties during the real-life phase. Only a minority moved from progressing with difficulties to progressing normatively, and equally many deteriorated during follow-up. Improved functioning as a consequence of alleviating gender dysphoria and passing better in the desired role is commonly assumed but has not previously been researched in relation to education/work. Our findings suggest that treatment of gender dysphoria does not suffice to improve functioning in education and working life. Difficulties in school adjustment and learning are common among gender-referred adolescents and often not properly addressed, on the assumption that treatment of gender dysphoria would relieve an array of problems [11,29]. Educational difficulties need to be fully addressed during adolescence regardless of gender identity.

On their developmental path towards emotional, social and economic independence from parents, adolescents gain competence in taking care of increasingly demanding matters outside home. Delays in this could be associated with gender dysphoria through psychiatric symptoms secondary to gender dysphoria and lack of self-confidence related to challenges in self-presentation. Such problems could be expected to be alleviated with gender affirming hormonal treatments. In taking care of matters at an age-appropriate level, a greater share had improved than had declined during the real-life phase. Thus, favourable progression of adolescent development was seen in the group studied, even if a fifth of the subjects continued to function on a lower than age-appropriate level during the real-life phase.

Need for treatment due to depression, anxiety and suicidality/self-harm was recorded less frequently during the real-life phase than before it. This is in line with the conclusion of a relatively recent meta-analysis [30] that in adults with gender dysphoria, cross-sex hormonal treatment alleviates anxiety, and may also reduce depression or depressive symptoms. However, need for psychiatric treatment overall did

not decrease from the level before and during the gender identity assessment to the real-life phase. New needs had also emerged about as frequently as need for treatment diminished. Cross-sex hormonal treatment is not enough to alleviate psychiatric comorbidities which in adolescents with gender dysphoria may also precede gender identity concerns [11] and will likely have equally many and complex underpinnings as they have in any population. A large-scale register study among adults likewise found that psychiatric needs were not alleviated with gender reassignment [12]. Depression, anxiety and suicidality/self-harm are often assumed to be secondary to gender dysphoria, and our findings may be interpreted as lending some support to that assumption among adolescents, similarly as earlier research seems to imply for adults [30].

Both earlier and concurrent need for psychiatric treatment were associated with not progressing age-appropriately at school/work and in taking care of matters outside home during the real-life phase, even though need for psychiatric treatment was, somewhat unexpectedly, not associated with functioning in peer relationships and romantic relationships. This further underlines the need to actively address psychiatric comorbidities among adolescents with gender dysphoria.

The study was based on file information on all adolescents diagnosed with transsexualism and proceeding to cross-sex hormone treatment after entering gender identity earlier than age 18 in one of the two centralized gender identity service facilities for minors in Finland. The two gender identity units for minors operate on similar principles, receive equal numbers of referrals and during the study period prescribed cross-sex hormones to similar numbers of adolescents. The follow-up period was approximately only a year, which inhibits drawing conclusions on long-term outcomes. However, as during adolescence, both physical, cognitive, emotional and social aspects of development are in a constant state of change [3], one year is a very relevant period.

Collected from medical files, the data is as accurate as clinical documentation can be. Because gender identity assessments and medical gender reassignments in minors involve numerous controversies (Kaltiala-Heino et al. [10]), the documentation is likely to be done particularly meticulously. The study unit operates within the field of adolescent psychiatry, and particular attention is always paid to adolescent development illustrated in age-appropriate functioning. Data collection was carried out in a structured way, which adds to the reliability of the study. Most of the recorded issues are clear-cut and concrete (living arrangements; progressing one class per year at school or having a job; socializing with peers in leisure time; being in (an offline) steady relationship). Age-appropriate capacity taking care of matters outside home may be somewhat more abstract and difficult to quantify. Ambiguous details were discussed between the authors and rated in consensus.

The disorders that were the reason for need for psychiatric treatment were recorded as they appeared in the documentation produced by the gender identity team or were recorded in case files obtained by the gender identity team from the adolescent's local services and classified on a robust level. They were not always systematically recorded with ICD-codes, and we were not able to ascertain the accuracy of the diagnostic work. However, the diagnoses mentioned in this paper represent problem categories that are the basis for treatment offered. A better understanding of psychiatric comorbidities could have been obtained by using structured diagnostic interviews.

Conclusion

Among adolescents diagnosed with transsexualism, difficulties in adolescent development and functioning in life domains appropriate to late adolescence do not disappear with cross-sex hormone treatment. Cross-sex hormone treatment may alleviate depression and anxiety but does not have a positive impact on psychiatric comorbidities at large. Even deterioration as regards psychiatric treatment needs and functioning occurs during the first year of cross-sex hormone treatment. Not all psychiatric and psychosocial problems in adolescents displaying gender dysphoria are secondary to gender identity issues and will not be relieved by medical gender reassignment. An adolescent's gender identity concerns must not become a reason for failure to address all her/his other relevant problems in the usual way.

Disclosure statement

No potential conflict of interest was reported by the authors.

Notes on contributors

Riittakerttu Kaltiala, MD, PhD, BSc, is professor of adolescent psychiatry in Tampere University and chief psychiatrist in adolescent psychiatry in Tampere University Hospital.

Elias Heino, cand med, studies medicine in Tampere University and participates in research on adolescent gender identity issues.

Marja Työläjärvi, MD, is adolescent psychiatrist in Tampere University Hospital. Her work focuses on adolescent gender identity and forensic issues.

Laura Suomalainen, MD, PhD, is chief adolescent psychiatrist in the adolescent gender identity service in Helsinki University Hospital.

ORCID

Riittakerttu Kaltiala http://orcid.org/0000-0002-2783-3892

References

[1] Blos P. The adolescent passage: developmental issues. New York: International Universities Press; 1979.
[2] Steinberg L. Cognitive and affective development in adolescence. Trends Cogn Sci. 2005;9:69–74.
[3] Paus T, Keshavan M, Giedd JN. Why do many psychiatric disorders emerge during adolescence? Nat Rev Neurosci. 2008;9:947–957.
[4] Havighurst RJ. Deelopmental tasks and education. Chicago (IL): University of Chicago Press; 1948.

[5] Palombo J, Bendicksen HK, Koch BJ. Guide to psychoanalytic developmental theories. New York: Springer; 2009.
[6] Erikson EH. Childhood and society. New York: W. W. Norton; 1950.
[7] American Psychiatric Association. Diagnostic and statistical manual of mental disorders. 5th ed. Washington (DC): American Psychiatric Association; 2013.
[8] World Health Organization. The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization; 1992.
[9] Seiffge-Krenke I, Gelhaar T. Does successful attainment of developmental tasks lead to happiness and success in later developmental tasks? A test of Havighurst's (1948) theses. J Adolesc. 2008;31:33–52.
[10] Kaltiala-Heino R, Bergman H, Tyolajarvi M, et al. Gender dysphoria in adolescence: current perspectives. AHMT. 2018;9:31–41.
[11] Kaltiala-Heino R, Sumia M, Tyolajarvi M, et al. Two years of gender identity service for minors: overrepresentation of natal girls with severe problems in adolescent development. Child Adolesc Psychiatry Ment Health. 2015;9:9.
[12] Dhejne C, Lichtenstein P, Boman M, et al. Long-term follow-up of transsexual persons undergoing sex reassignment surgery: cohort study in Sweden. PLoS One. 2011;6:e16885.
[13] Wood H, Sasaki S, Bradley SJ, et al. Patterns of referral to a gender identity service for children and adolescents (1976-2011): age, sex ratio, and sexual orientation. J Sex Marital Ther. 2013;39:1–6.
[14] Aitken M, Steensma TD, Blanchard R, et al. Evidence for an altered sex ratio in clinic-referred adolescents with gender dysphoria. J Sex Med. 2015;12:756–763.
[15] Kaltiala-Heino R, Työläjärvi M, Lindberg N. Gender dysphoria in adolescent population: a 5-year replication study. Clin child psychol Psychiatry. 2019;24:379–387.
[16] Kaltiala-Heino R, Lindberg N. Gender identities in adolescent population: methodological issues and prevalence across age groups. Eur Psychiatry. 2019;55:61–66.
[17] Cohen-Kettenis PT, Klink D. Adolescents with gender dysphoria. Baillieres Best Pract Res Clin Endocrinol Metab. 2015;29:485–495.
[18] Coleman E, Bockting W, Botzer M, et al. The Standards of Care of the World Professional Association for Transgender Health, 7th version. Int J Transgend. 2012;13:165–232.
[19] Costa R, Dunsford M, Skagerberg E, et al. Psychological support, puberty suppression, and psychosocial functioning in adolescents with gender dysphoria. J Sex Med. 2015;12:2206–2214.
[20] de Vries AL, Steensma TD, Doreleijers TA, et al. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. J Sex Med. 2011;8:2276–2283.
[21] Chew D, Anderson J, Williams K, et al. Hormonal treatment in young people with gender dysphoria: a systematic review. Pediatrics. 2018;141:e20173742.
[22] Heneghan C, Mahtani K BMJ. Gender-affirming hormone in children and adolescents – evidence review. 2019 [cited 2019 Feb 25]. Available from: https://blogs.bmj.com/bmjebmspotlight/2019/02/25/gender-affirming-hormone-in-children-and-adolescents-evidence-review/
[23] Tilastokeskus a. Perheet 2014 [cited 2019 Feb 25]. Available from: https://www.tilastokeskus.fi/til/perh/2014/02/perh_2014_02_2015-11-27_tie_001_fi.html
[24] Nikander T. Nuoret muuttavat omilleen yhä nuorempina. 2009 [cited 2019 Feb 25]. Available from: https://www.stat.fi/artikkelit/2009/art_2009-03-16_004.html?s=0
[25] Mayer KH, Garofalo R, Makadon HJ. Promoting the successful development of sexual and gender minority youths. Am J Public Health. 2014;104:976–981.
[26] de Vries ALC, Steensma TD, Cohen-Kettenis PT, et al. Poor peer relations predict parent- and self-reported behavioral and emotional problems of adolescents with gender dysphoria: a cross-national, cross-clinic comparative analysis. Eur Child Adolesc Psychiatry. 2016;25:579–588.
[27] Bungener SL, Steensma TD, Cohen-Kettenis PT, et al. Sexual and romantic experiences of transgender youth before gender-affirmative treatment. Pediatrics. 2017;139:e20162283.
[28] Kaltiala-Heino R, Työläjärvi M, Lindberg N. Sexual experiences of clinically referred adolescents with features of gender dysphoria. Clin Child Psychol Psychiatry. 2019;24:365–378.
[29] Holt V, Skagerberg E, Dunsford M. Young people with features of gender dysphoria: demographics and associated difficulties. Clin Child Psychol Psychiatry. 2016;21:108–118.
[30] Costa M, Colizzi R. The effect of cross-sex hormonal treatment on gender dysphoria individuals' mental health: a systematic review. Neuropsychiatr Dis Treat. 2016;12:1953–1966.

Kaltiala-Heino et al. Child and Adolescent Psychiatry and Mental Health (2015) 9:9
DOI 10.1186/s13034-015-0042-y

CHILD & ADOLESCENT PSYCHIATRY & MENTAL HEALTH

RESEARCH Open Access

Two years of gender identity service for minors: overrepresentation of natal girls with severe problems in adolescent development

Riittakerttu Kaltiala-Heino[1,2*], Maria Sumia[2], Marja Työläjärvi[2] and Nina Lindberg[3,4]

Abstract

Background: Increasing numbers of adolescents present in adolescent gender identity services, desiring sex reassignment (SR). The aim of this study is to describe the adolescent applicants for legal and medical sex reassignment during the first two years of adolescent gender identity team in Finland, in terms of sociodemographic, psychiatric and gender identity related factors and adolescent development.

Methods: Structured quantitative retrospective chart review and qualitative analysis of case files of all adolescent SR applicants who entered the assessment by the end of 2013.

Results: The number of referrals exceeded expectations in light of epidemiological knowledge. Natal girls were markedly overrepresented among applicants. Severe psychopathology preceding onset of gender dysphoria was common. Autism spectrum problems were very common.

Conclusion: The findings do not fit the commonly accepted image of a gender dysphoric minor. Treatment guidelines need to consider gender dysphoria in minors in the context of severe psychopathology and developmental difficulties.

Keywords: Transsexualism, Gender dysphoria, Sex reassignment, Adolescent development

Introduction

According to the ICD-10 [1], transsexualism involves a desire to live and be accepted as a member of the opposite sex, usually accompanied by the wish to make one's body as congruent as possible with one's preferred sex through surgery and hormonal treatment. The desire has to be persistent and not a symptom of a mental disorder. Gender dysphoria refers to dysphoria experienced due to the incongruence between a person's inner perception of her/his gender, and the incongruous bodily reality. The term Gender Dysphoria has also recently been adopted as the diagnostic category in DSM-5 [2]. Psychotherapeutic approaches have not proven successful in relieving gender dysphoria, and social, juridical, medical and surgical sex reassignment (SR) is nowadays the treatment of choice [3]. Sex reassignment with hormonal and/or surgical treatments has been reported to improve social, psychological and sexual well-being and functioning.

Surveys based on the Child Behaviour Checklist [4] report that 2-5% of children aged up to seven, as reported by their parents, "behaves like opposite sex" and 1-2% "wishes to be of opposite sex", but cultural issues likely play a major role in whether a child's behavior is perceived as gender atypical. Consultations due to gender identity are generally more often sought for boys than girls, which may suggest greater gender variation in boys, but also that effeminate behaviours in boys are perceived as more of a problem than tom-boyishness in girls [5,6].

Of children with even severe gender dysphoria and cross-sex identification, about 85% do not develop a persistent transsexual identity in adolescence (reviewed in [7]). Reliable indicators are not so far available regarding which gender dysphoric children cease to be so in puberty and who develop transsexual identity [8]. Medical interventions are therefore not warranted in pre-

* Correspondence: merihe@uta.fi
[1]University of Tampere, School of Medicine, 33014 University of Tampere, Tampere, Finland
[2]Department of Adolescent Psychiatry, Tampere University Hospital, 33380 Pitkäniemi, Finland
Full list of author information is available at the end of the article

pubertal children. In light of current knowledge, transsexual identity in adolescence is persistent and medical interventions may be appropriate. According to the treatment model developed in the Netherlands (Dutch model), early treatment may include delaying puberty after its first stages with GnRh analogues, and administering cross-sex hormones from about age 16 [9,10]. This approach is recommended when childhood gender dysphoria exacerbates in puberty, no (primary) severe psychopathology is present, and the young person has appropriate developmental support and support for the process from her/his primary caregivers (parents). The rationale with GnRh analogue treatment is to prevent the undesired development of secondary sex characteristics and thereby facilitating later transition to the desired role, and postponing complicated and irreversible treatment decisions to a more mature age. Psychopathology largely attributed secondary to gender dysphoria is expected to be relieved by puberty blocking and resolved by sex reassignment [5,11,12].

In the past decade, the numbers of referrals to child and adolescent gender identity services have been on the increase across Europe (personal communications in 2013 and 2014 from UK, NL, Spain, Sweden child and adolescent gender identity teams) and in Canada [13]. It is not known whether this represents a true increase in gender dysphoria, lowered thresholds for seeking help for it or perhaps cultural developments that promote the conceptualization of developmental challenges as being rooted in sex and gender.

In Finland, the legislation stipulates that a transsexual person may be recognized in law as a member of the desired sex and have access to hormonal and surgical sex reassignment (in public health care) (act 2002/563). A psychiatric assessment by a specialized gender identity team is a prerequisite for legal as well as surgical sex reassignment, both of which have a lower age limit of 18. The specialized psychiatric assessment by a gender identity team is centralized to two university hospital psychiatric clinics, Tampere and Helsinki University Hospitals, in the country (codes 1053/2002 and 476/2010).

Since 2011, specialized adolescent psychiatric gender identity teams have been available for minors at the above mentioned two university hospitals. The excessive number of referrals, exceptional sex ratio and severity of general psychopathology among the applicants compared to what might have been anticipated on the basis of the literature called for clinical attention from the beginning of the service. The aim of this study is to describe the adolescent applicants for legal and medical sex reassignment in terms of sociodemographic, psychiatric and gender identity related factors and adolescent development in order to initiate a scientific discussion on the meaning of these observations.

Materials and methods

The study comprises a retrospective chart review of all the SR applicants attending for assessment by one of the two adolescent gender identity services in Finland (Tampere University Hospital, Department of Adolescent Psychiatry) in 2011–2013. Altogether 49 adolescents were referred to assessment for sex reassignment and invited to their first meeting during the study periods, but two adolescents declined to start the evaluation. Thus 47 adolescents are included in this study. Of these, one was mutistic and did not provide any information; for this young person, information on personal experiences is missing but information from case records and parents could be used.

The assessments take place in an outpatient setting and comprise structured and free format assessments and interviews with an adolescent psychiatrist, a psychiatric nurse, a social worker and a psychologist. The adolescent and her/his parents/guardians are seen together and separately by all the multi-disciplinary team members. Psychiatric and medical files are requested from all previous health care contacts of the adolescent, with due permission from her/him and her/his parents. After completing all the assessments, the multi-disciplinary team discusses the diagnosis as to gender identity and mental disorders, eligibility for hormonal SR treatments and possible other needs to be met and recommendations to be given regarding gender identity needs and mental health needs when appropriate. All the below described measures were collected using all the material available after the assessment. The study received approval from the ethics committee of Pirkanmaa Hospital District.

Measures

Sociodemographic variables collected were age, natal sex, family structure (living with both parents/one parent/neither parent) and parental education (professional/intermediate/skilled non-manual/skilled manual/unskilled or not in employment). Further background information included the reason for referral (sex reassignment, definite wish/sex reassignment, possible reassignment/other) and parental homosexuality or transsexualism (yes/no).

Throughout the discussion of our own research we use the terms “gender dysphoria” and “gender dysphoric” to refer to the experienced gender incongruence among our applicants, regardless of whether they fulfill the diagnostic criteria for Gender Dysphoria in the DSM-5. For the present study we recorded whether there had been signs of gender dysphoria/gender incongruence in childhood (before age 12) (Table 1). Age of onset of conscious gender concerns and age when the applicant was convinced that s/he is transsexual were recorded. If the adolescent was already living in the desired role (Table 1), it was recorded for how long.

Table 1 Variable descriptions for childhood gender dysphoria, bullying and social isolation

Gender dysphoria/gender incongruence in childhood (<12 years of age)	
Childhood gender dysphoria/incongruence present	Childhood gender dysphoria/incongruence not present
• explicit gender dysphoria or marked and persistent cross-gender identification on behavioural level even without explicit verbalization of one's gender related thoughts and feelings in childhood	• no signs of gender dysphoria/incongruence in childhood
Gender presentation, living in desired role	
Classified as living in the desired role	Not classified as living in the desired role
• the applicant had officially changed her/his name to a gender neutral one or arranged his/her registration in school, and being consistently called, by a name suggesting the desired sex; being always presented to new people as being of the desired sex; being treated by family, teachers/ employers, friends, schoolmates as well as by new people as a member of the desired sex	• the adolescent had not made any attempt to live and be treated in the desired role
	• the adolescent dressed gender neutrally and asked the family to use a name indicative of the desired sex, but was actually not living in any social role outside the family due to isolation from social interactions
• some of these young people had explicitly "come out" in school and openly made a transition to the desired role; some, with the support of some community key adults, had adults, had totally concealed their natal sex from the school/workplace	• some of the adolescents in this group were almost totally isolated in their homes (not going to school or work, not meeting peers), some attended school but were isolated from social interactions there and elsewhere
Bullying	
Significantly subjected to bullying	Not subjected to bullying
• the applicant and/or her/his parents considered that there had been significant and traumatic victimization.	• no recollection of being bullied
	• if ever bullied, the adolescent described it as non-significant ("maybe sometimes", "not more than anyone else").
• a) related to gender presentation or sexual identity: name-calling, spreading rumours and the like related to gender presentation/sexual identity	
• b) not related to gender or sexual identity: bullying was related to other issues like weight, interests, belonging or not belonging to a certain group etc.	
Isolation	
Periods of isolation	No isolation
• periods of not having contact with peers outside of arranged study-related activities at school - or not even that, if not attending school	• no interruptions in attending age appropriate daily programme (usually school), having age-appropriate contacts with peers
• having same-age contacts only with one's own siblings	
• keeping (tenuous, infrequent) contact with one or two peers only despite previously having been normatively engaged with peers	
• contacts outside the family only via Internet	

Previous and current psychiatric history was recorded. Previous files were not always complete, and diagnoses were not always accurately defined in terms of ICD-10 diagnostic codes. Thus, we recorded 1) whether the young person had been in contact with psychiatric services prior to entering the gender identity service (yes/no), 2) whether the previous contact had been because of gender concerns or psychiatric symptoms (gender issues only/psychiatric symptoms only/both), 3) what kind of problems the young person had displayed (anxiety, depression, suicidal behaviours, conduct problems, autism spectrum related problems, substance abuse, psychotic symptoms, other; all recorded yes/no), and 4) the temporal relationship between psychiatric symptoms and gender dysphoria/identity concerns (psychiatric symptoms emerged earlier/gender dysphoria and gender identity concerns emerged first).

Peer-related difficulties were recorded being subjected to bullying at school (yes/no) and isolation from peers (yes/no) (Table 1). Of bullying it was recorded whether it happened before, after or both before and after of the onset of gender dysphoria, and whether it was related to gender presentation or sexual orientation. Of social isolation it was recorded whether it occurred before, after or both before and after the onset of gender dysphoria.

Statistical analyses

All the variables were recorded in a structured form developed for this research. Descriptive analysis was conducted using statistical methods for quantitative data. We report frequencies and means (sd) where appropriate. Between groups comparisons are made with cross-

tabulations and chi-square statistics/Fisher's exact test, and with t-test where appropriate.

Qualitative observations

The qualitative content analysis approach [14] was applied to illustrate, based on all material recorded in case histories, different groups of gender dysphoric adolescents, or different developmental pathways resulting in the adolescent now perceiving the need to apply for sex reassignment. This was carried out by condensing and extracting from all material recorded in the case histories similar and different developmental patterns and descriptions of experiences that could be used to create mutually exclusive model stories, or trajectories that would include all the studied adolescents and not allow for assigning a given adolescent to more than one trajectory. The model stories were not defined in advance but they were formed in a data-driven process, the outcome of which is presented.

Results

Demographics

Of the applicants included in the present study, 41 were natal girls and 6 were natal boys. Their mean age (sd) at entering assessment was 16.04 (0.57) years for natal boys and 16.66 (1.07) for natal girls (p = 0.18). Of these, 49% (23) were living with both their biological parents, 39% (18) with one biological parent, and 13 % (6) in child welfare foster placements or independently. Parental education was distributed as follows: 16% (8) professional, 5% (2) intermediate, 22% (10) skilled non-manual, 43% (20) skilled manual, and 14% (7) were unskilled or not participating in work life. None of the applicants had transsexual or homosexual parents.

Gender dysphoria

Of the applicants, 32% (14/47) reported having started to consciously question their gender before age 12, 62% (30/47) at 12 or later, and three applicants (6%) could not define this. Most commonly (one in five) these concerns had started at age 14. There were altogether five applicants (11%) who during childhood had persistently expressed gender dysphoria and/or identified with the opposite sex, and three (6%) who during childhood had transiently displayed gender dysphoria and a desire to be of the opposite sex. A further nine applicants (19%) had been tomboyish girls but had not questioned their gender or experienced dysphoria, and as to most of the applicants (30/47, 64%), neither the young person nor her/his parents recalled gender dysphoria or cross-gender behaviors during childhood.

During the assessment process, 72% (34/47) of the applicants were sure about feeling they were of the opposite sex to their natal and about pursuing sex reassignment, but 28% (13/47) were not sure about their feelings regarding gender identity and/or sex reassignment. There was no difference between natal girls and natal boys in this regard. Of those who felt sure about their cross-gender identity, 15% (5/34) recalled reaching the conclusion before age 12, 79% (27/34) at 12 or later, and two (6%) could not define at what age they had reached the conclusion. There was no difference between natal girls and natal boys. The time frame from first becoming aware of gender dysphoria to being sure of one's own cross-gender identity ranged from 0 to 7 years, with mean 1.6 (sd 2.1) years.

Of all the applicants, 38% (17/47) were living in the desired role when the assessment was completed, 50% (3/6) of the natal boys and 37% (15/41) of the natal girls (p = 0.41). Of those applicants who expressed certainty about being of other than their natal sex and desiring physical and legal sex reassignment, 47% (16/34) were living in the desired role. Of those who were living in the desired role, the mean (sd)/median time of living in the role was 28.3 (17.9)/24.0 months for natal boys, and 29.8 (39.2)/12 months for natal girls (p = ns).

Peer relationship difficulties

Of the applicants, 57% (27/47) had been significantly bullied at school, 53% (25/47) in primary school (grades 1–6, ages 7–13 yrs) and 45% (21/47) in secondary school (grades 7–9, ages 13–16 yrs). Of those who had been victims of bullying, 73% (19/27) had been bullied before they came to think about their gender identity, 8% (2/27) after starting to think about gender issues, and 19% (5/27) both before and after. Of those bullied, 27% (7/26) reported that bullying had been related to gender presentation or sexual identity, and 73% (19/26) had been bullied due to some other reasons (see Table 1).

Natal girls and natal boys had been bullied equally frequently. Natal girls tended more often to report having been bullied only before the onset of gender dysphoria, and natal boys more often both before and after the onset of gender dysphoria (girls: 78% (17/23) only before, 9% (2/23) only after, 13% (3/23) before and after vs. boys: 33% (1/3) only before, none only after, 67% (2/3) both before and after, p = 0.08). Among natal boys gender presentation and/or sexual identity had always been the topic of the bullying, among natal girls 83% (19/23) had been bullied for something else and 17% (4/23) due to gender presentation/sexual identity (p = 0.01).

Of the applicants, 45% (21/47) had presented with periods of isolation from peer relationships; 32% (15/47) before and 40% (19/47) after the onset of gender dysphoria, and 43% (20/47) were socially isolated during the SR assessment. Twenty-eight per cent (13/47) were isolated in all three observed periods. Social isolation was equally common among natal boys and girls applicants.

Psychiatric treatment and psychopathology

Seventy-five per cent of the applicants (35/47) had been or were currently undergoing child and adolescent psychiatric treatment for reasons other than gender dysphoria when they sought referral to SR assessment, and two more were contacted with general adolescent psychiatric services soon after entering the SR assessment. Sixty-four per cent (30/47) were having or had had treatment contact due to depression, 55% (26/47) due to anxiety disorders, 53% (25/47) due to suicidal and self-harming behaviours, 13% due to psychotic symptoms (6/47), 9% (4/47) due to conduct disorders, 4% (2/47) due to substance abuse, 26% (12/47) due to autism spectrum disorder, and 11% (5/47) due to ADHD. One severe case of anorexia nervosa was noted. Of the applicants, 68% (32/47) had had their first contact with psychiatric services due to other reasons than gender identity issues. Natal boys and natal girls had equally commonly been treated for psychiatric disorders except for ADHD which had been more commonly treated in natal boys (50% vs.5%, $p = 0.01$). The mean number of distinct psychiatric problems was 2.3 (sd 1.7), with no difference between natal girls and natal boys.

The different groups

Five different mutually exclusive groups (a - e below) were identified that differed as to onset of gender dysphoria and cross-gender identification, psychopathology and adjustment/difficulties in social relationships, and the temporal relationships between these. They are presented in Table 2.

We carried out logistic regression analyses to detect what kind of presenting features were associated with belonging to the last, confused group of adolescents with gender dysphoria (e) when entered in the model simultaneously. This was appropriate because psychiatric symptoms and psychosocial functioning are strongly interrelated. Age and natal sex were not predictive of belonging to the confused group. Each psychiatric problem, being subjected to bullying, presenting with periods of isolation, number of different psychiatric problems, and months living in desired role were each in turn entered as independent variables, controlling for age and natal sex. When controlling for age and natal sex, group memberships was predicted by anxiety (OR 4.8, 95% CI 1.4-17.0), suicidality (OR 5.7, 95% CI 1.7-20.3), number of different psychiatric symptoms (OR

Table 2 The different groups of gender dysphoric adolescents seeking SR

Early onset gender dysphoria, exacerbates in puberty	
a) with no with no significant psychopathology and developmental problems (n = 2)	b) with considerable psychopathology and developmental problems (n = 3)
• very mild or no psychopathology across childhood and until the assessment	• severe psychopathology that had previously and currently required specialist level child and adolescent psychiatric care (autism spectrum disorder, OCD, Tourette, anorexia nervosa, suspected psychotic episodes or psychosis high risk, specific learning difficulties)
Adolescent onset gender dysphoria, where transsexual identity appeared established	
c) without, or with only mild psychopathology and developmental difficulties (n = 10)	d) with severe psychopathology and developmental difficulties (n = 9)
• mild to moderate depression or anxiety, could be considered secondary to gender dysphoria, or was transient, and did not impair functioning in social relationships or school	• psychiatric problems that warranted specialist level adolescent psychiatric treatment, either in treatment at the beginning of their SR assessment, or treatment contact was arranged during the SR assessment
• age-appropriate social relationships and leisure time activities, participation in age-appropriate educational activities (comprehensive, vocational or upper secondary school)	• autism spectrum disorders (3), major depression (3), social phobia (5), substance abuse problems (1) or a history of conduct disorder and trauma (2) (several had two disorders); clearly more severe psychopathology than what was seen in group c
e) Adolescent onset gender dysphoria, identity confused development (n = 23)	
• In childhood, no gender dysphoria nor cross-gender behaviors	
• For most of their primary school years (age 7–12 years) felt excluded	
• Persistent experiences of bullying before the onset of gender dysphoria	
• In adolescence, social anxiety and depression, most often with self-harm and suicidal preoccupation if not suicide attempts	
• Isolated	
• Long periods of not attending school, or if attended school, did not engage in peer contacts outside learning situations arranged by teachers.	
• Did not meet with same-aged peers in leisure time, or they met with few peers and only if their parents arranged it; many in contact only with their family members.	
• Socially and/or academically marginalized	
• Very high expectations that SR would solve their problems in social, academic, occupational and mental health domains	

1.7, 95% CI 1.1-2.6), and presenting with periods of isolation (OR 9.0, 95% CI 2.3-34.7). However, when presenting with periods of social isolation was entered into any other model, the other independent variables were leveled out, suggesting that social isolation was the strongest factor predicting membership of the problematic, identity confused group.

Discussion

The number of referrals exceeded expectations. Given the most cited epidemiological figures among adults, 1:10 000–1:30000 MtF and 1:40 000–1:100 000 FtM [6], in Finnish population, 6–18 boy-to-girl adolescents and 2–4 girl-to-boy adolescents aged 13–18 would be expected. The number of referrals to the study unit already doubled the less conservative estimates based on adult figures. Referrals to the other adolescent gender identity unit amount to equal numbers, and the natal girl:boy ratio in referrals is also similar in the other unit (Tainio V-M, personal communication). Valid epidemiological research on incidence and prevalence of transsexualism or gender dysphoria at large among adolescents is not available [6]. The adult figures cited above are based on treatment seeking, as are the numbers presented in the present study. Gender dysphoria may be more common among adolescents than among adults, or it may be increasing in younger age cohorts.

Not all applicants could be seen as presenting with established transsexual identity, even though they suffered gender dysphoria. Excluding the confused (e) group in our data, 3 boy-to-girl and 21 girl-to-boy applicants were identified who displayed transsexual identity that appeared established, unique, and not part of more general identity confusion or secondary to severe mental disorders. Given that these numbers are based on half of the adolescent gender identity assessments in Finland, the findings further suggest that severe and persistent gender dysphoria/transsexualism in adolescence may be more common than hitherto assumed.

The natal girl:boy ratio among the adolescent SR applicants was very high. In prepubertal children referred to gender identity services, boy:girl ratio is reportedly 3–6:1, with some variation across countries presumably due to cultural reasons [5,13]. Previously a more even boy:girl ratio has been suggested in adolescents seeking sex reassignment than among child samples [13], and a recent paper from Germany reported natal boy:natal girl ratio of 0.81 among 268 minors diagnosed with gender identity disorder from 1987–2013 [15]. Among adults, there seems to be remarkable variation across countries in the ratio of natal men:natal women seeking for sex reassignment [16]. In Western countries natal male transsexuals exceed natal females transsexuals. A German study demonstrated that the natal male:natal female ratio among transsexual people has changed to more equal towards 2000's that what it was in earlier decades [16]. However, the overrepresentation of girls on our sample differs still from these more recent trends, and it is similar in both the two Finnish centers. We have so far no explanation for this great overrepresentation of natal girls seen in our material, and equalizing of sex ratio demonstrated by others [13,15,16]. Cultural trends may somehow influence this. May be more permissive societal attitudes allow "coming out" as gender variant more easily than before. However, why this would concern primarily girls remains an open question.

Of children and adolescents, 10-15% are regularly (weekly) involved in school bullying [17]. Of the adolescent SR applicants, more than a half had been subjected to bullying. Even if in the present study it was not possible to verify exactly how frequently the applicants had been bullied, we only recorded bullying that the adolescent and her/his parents perceived as significant: particularly intensive, vicious, long-term and traumatizing. However, in more than two thirds of the cases, bullying had occurred before the onset of gender dysphoria, and was not targeted at gender or sexual identity. Bullying is an unspecific risk factor for developmental problems rather than a problem specifically related to gender identity. That natal boys were more commonly bullied because of gender presentation suggests that effeminate characteristics in boys are less tolerated than masculine self-presentation in girls.

Peer relationships are of the outmost importance during adolescent development [18-20], and social isolation from peer relationships suggests developmental difficulties and impaired mental health [21-24]. In the present sample, isolation was extremely common and also the strongest predictor of membership of the "confused" group.

More than three quarters of the adolescent SR applicants had needed and/or currently needed specialist level child and adolescent psychiatric services due to psychiatric problems other than gender dysphoria. Specialist level child and adolescent psychiatric services are provided exclusively for severe disorders in Finland [25,26]. The recorded comorbid disorders were thus severe and could seldom be considered secondary to gender dysphoria. This utterly contradicts the findings in the Dutch child and adolescent gender identity service, where two thirds of adolescent SR applicants did not have psychiatric comorbidity [27]. In a recent German study, 43% of children and adolescents seen in a specific gender identity unit suffered from major psychopathology [15]. For the time being, we are unable to explain why Finnish adolescent SR applicants appear psychiatrically much more disturbed than has been reported elsewhere, but our findings warrant attention. The

treatment guidelines for adolescent gender dysphoria may require extensions taking into account the needs of those with severe psychopathology and identity confusion, very unlikely currently eligible to medical SR.

The overlap between autism spectrum disorders and gender dysphoria has been recognized before [28]. In a Dutch gender identity service, 9.4% of adolescents presented with autism spectrum disorder. In our sample, 26% of the adolescent SR applicants were diagnosed to be on the autism spectrum. These diagnoses had mainly been made during the adolescents' previous psychiatric treatment in our hospital or elsewhere, but three such diagnoses were also made by our team. In our hospital, the ADOS [29] is used with the minors, and the 3Di [30] or ADI-R [31] with parents to diagnose autism spectrum disorders. We could not systematically review with which protocols the diagnoses had been made elsewhere in the country, but in our clinical opinion there was no reason to doubt them. It is currently not known why autism spectrum is overrepresented in gender dysphoric children and adolescents. The conditions could be truly co-occurring. Prenatal exposure to high levels of testosterone could be involved in the development of both conditions, especially for girls with autism spectrum disorder, but this leaves the comorbidity in males unexplained. Gender identity issues could arise from autism spectrum people's predisposition toward unusual interests, or gender dysphoria in ASD could represent OCD rather than genuine gender identity issues. The cross-gender behaviour in ASD minors could also rather represent non-normative sexual interests or unusual sensory preferences [28]. Our clinical impression is that a long-standing feeling of being different and an outsider among peers could play a role in ASD children developing gender dysphoria in adolescence. In our clinical sample of gender dysphoric adolescents, autism spectrum disorders by far exceeded the prevalence of 6/1000 suggested for general population [32], and almost three-fold that in the sample of deVries et al. [28]. Autism spectrum needs to be taken seriously in considering treatment guidelines for child and adolescent gender dysphoria. Given the nature of ASD, particularly ASD children's and adolescents' difficulties in adjusting to changes, profound changes in their own bodies with SR treatments may pose a major challenge to psychological adjustment, and ASD adolescents may be particularly rigidly unwilling to consider this in advance.

In the international literature on gender dysphoria in minors, the most often portrayed picture is that of childhood cross-gender identification/gender dysphoria, where gender dysphoria exacerbates in puberty due to the development of secondary sex characteristics. Our findings suggest that there are many more developmental pathways that may also need different treatment approaches. In our data, most of the adolescents first presented with gender dysphoria and cross-gender identification well after the onset of puberty, and the vast majority suffered significant psychopathology and broader identity confusion than gender identity issues alone. It is important to be able to openly discuss these alternative presentations of gender dysphoria in order to find appropriate treatment options.

Adolescence is a period of identity formation. From early to late adolescence, identity develops from fragmented and contextual identity experience to endogenous, permanent and integral identity that remains constant across contexts and interactions [33]. Identity is formed through diverse physical and psychological developments and in relation to other people and the social environment [34,35]. An adolescent also faces fundamental identity challenges in the domains of religion, worldview, ethnicity, sexuality and the like. Identification with various groups is often passionate during adolescence, but the object of identification may also change, even several times [34-37]. Adolescents are more suggestible and submit more readily to group pressure to gain acceptance [38]. Adolescence is a period of maturation of social cognition, and a prerequisite for the maturation of social cognition is the maturation of the central nervous system that continues to the third decade of life [39]. During puberty and adolescent development there may be some overlap between normative testing of sexuality and gender roles in the one end, and gender dysphoria as a disorder in the other end of the spectrum. This would implicate that GD in adults and in adolescence may not be the same issue in general. For these reasons it is more challenging to assess whether the gender identity of an adolescent is so firmly established that physical intervention is indicated than it is to assess this among adults.

In the majority of the applicants, gender dysphoria presented in the context of wider identity confusion, severe psychopathology and considerable challenges in the adolescent development. At this point it is not possible to predict how gender dysphoria in this group will develop: will gender dysphoria in these adolescents cease with the resolution of wider developmental problems, or perhaps consolidate later into transsexual identity, with the completion of the developmental tasks of adolescence.

Methodological considerations

The present paper is based on information on all adolescents who entered the assessment for sex reassignment in Finland in 2011–2013 by one of the two centralized adolescent gender identity teams in the country. The basis for choosing one or another of the two centers was geographical and not likely to create bias due to subject selection. It is further known that number of referrals

during the study period well as natal girl:natal boy ratio are similar in both centers.

The data collection was systematic and structured, which adds to the reliability of the findings. The data collection took place in the form of retrospective chart review of files created during a comprehensive assessment period by a multi-disciplinary team. Thus data collection was unlikely to bias the assessments in any way. Comprehensive assessments by a multi-disciplinary team are likely to provide reliable and valid data. The multi-disciplinary team collected information from the applicants themselves, from their parents, from previous case histories and by their own psychometric measurements. The applicants themselves might be prone to interpret a variety of their problems as being a result of gender incongruence, even if the problems actually were independent of gender identity issues or even predisposing to gender incongruence. In this study we attempted to avoid bias due to subjects' interpretation by using multiple source of information.

However, the data is relatively small and does not permit complex analyses. The study remains descriptive and cannot shed light on causal relationships. Some information of interest for the research was occasionally missing in the files, because the files were primarily created for clinical purposes, not for research.

The validity of the diagnoses in previous psychiatric contacts needs to be considered with certain caution. Previous files were not always complete and did not provide diagnoses according to ICD-10, and we were not able to check in the databases of the previous treatment providers what ICD diagnoses were recorded there. Thus, we recorded reasons for previous treatment based on verbalizations in the referrals and available copies of previous files. This only allowed a rough descriptive classification to problems related to anxiety, depression, suicidal behaviours, conduct problems, autism spectrum related problems, substance abuse, psychotic symptoms and other. We only recorded these problems if the adolescent had had a psychiatric treatment contact. The data gives a picture of the primary problems in previous psychiatric treatment contacts but not of all possible symptoms. Thus, our figures for problems related to anxiety, depression etc. are likely underestimates. It was also not possible to obtain exact information of when the various symptoms and disorders had been present and for how long time, except for autism which is of course assumed a lifetime condition. However, as clinical research on adolescent SR applicants is scarce, descriptive studies are valuable in providing a basis for discussion and international comparisons that are needed in order to create optimal clinical treatment guidelines.

Psychotic symptoms in our data mainly comprise brief and limited hallucinatory experiences. Psychotic symptoms were recorded if there were descriptions of hallucinations in the files, or of the previous files mentioned "psychotic symptoms" even when not giving more detailed descriptions. However, none of the applicants had a diagnosis of schizophrenia or schizoaffective disorder. Assessing gender dysphoria in the context of schizophrenia spectrum psychoses would be inappropriate. Doctors/units primarily contacted would very unlikely refer a patient with schizophrenia or schizoaffective disorder in gender identity assessments. Current psychotic symptoms would result in the gender identity team promptly referring the young person to general adolescent psychiatric care.

The findings cannot be generalized to all adolescents experiencing gender variation. Not all gender incongruent people perceive a need to seek for SR, or find it timely during adolescence.

Conclusion

Adolescents seeking sex reassignment represent a variety of developmental pathways differentiated by the timing of onset of gender dysphoria, psychopathology and developmental difficulties. It is important to be aware of the different groups, or developmental pathways, in gender dysphoric adolescents in order to be able to find appropriate treatment options. In the presence of severe psychopathology and developmental difficulties, medical SR treatments may not be currently advisable. Treatment guidelines need to be reviewed extended to appreciate the complex situations.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

All authors participated in designing the present study and formulating the study questions. RKH and MS collected the data from case files. RKH performed the data analysis. All the authors participated in discussing the results and writing the manuscript. All authors read and approved the final manuscript.

Author details

[1]University of Tampere, School of Medicine, 33014 University of Tampere, Tampere, Finland. [2]Department of Adolescent Psychiatry, Tampere University Hospital, 33380 Pitkäniemi, Finland. [3]Faculty of Medicine, University of Helsinki, BOX 3300014 Helsinki University, Helsinki, Finland. [4]Forensic Psychiatry, Helsinki University Hospital, Helsinki, Finland.

Received: 28 November 2014 Accepted: 19 March 2015
Published online: 09 April 2015

References

1. World Health Association. International Statistical Classification of Diseases and Related Health Problems, 10th revision (ICD-10). Geneva: WHO; 1992.
2. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders, Fifth edition DSM-5). Washington CD: American Psychiatric Publishing; 2013.
3. Coleman E, Bockting W, Botzer M, Cohen-Kettenis P, DeCuypere G, Feldman J, et al. The Standards of Care of the World Professional Association for Transgender Health, 7th version. Int J Transgenderism. 2011;13:165–232.
4. Achenbach TM, Edelbrock C. Manual for the Child Behavior Checklist and Revised Child Behavior Profile. Burlington, VT: University of Vermont, Department of Psychiatry; 1983.

5. Möller B, Schreier H, Li A, Romer G. Gender identity disorder in children and adolescents. Curr Probl Pediatr Adolesc Health Care. 2009;39:117–43.
6. Zucker KJ, Lawrence AA. Epidemiology of gender identity disorder. Int J Transgenderism. 2009;11:8–18.
7. Steensma T. From gender variance to gender dysphoria. Psychosexual development of gender atypical children and adolescents. Academic dissertation, Vrije Universiteit Amsterdam. Ridderprint: Amsterdam, NL; 2013.
8. Steensma DT, Biemond R, deBoer F, Cohen-Kettenis PT. Desisting and persisting gender dysphoria after childhood: a qualitative follow-up study. Clin Child Psychol Psychiatry. 2011;16:499–516.
9. Cohen-Kettenis P, Steensma T, deVries AL. Treatment of adolescents with gender dysphoria in the Netherlands. Child Adol Psychiatric Clinics N Am. 2012;20:698–700.
10. deVries AL, Cohen-Kettenis P. Clinical management of gender dysphoria in children and adolescents. The Dutch approach. J Homosexuality. 2012;59:301–20.
11. Wren B. Early physical intervention for young people with atypical gender identity development. Clin Child Psychol Psychiatry. 2000;5:220–31.
12. deVries A, Cohen-Kettenis P, Delemarre-van de Maal H. Caring for transgender adolescents in BC: suggested guidelines. Clinical management of gender dysphoria in adolescents. Vancouver, Canada: Vancouver Coastal Health, Transcend Transgender Support & Education Society and the Canadian Rainbow Health Coalition; 2006.
13. Wood H, Sasaki S, Bradley S, Singh D, Fantus S, Owen-Anderson A, et al. Patterns of referral to a gender identity service for children and adolescents (1976–2011): age, sex ratio, and sexual orientation. J Sex Marital Ther. 2013;39:1–6.
14. Graneheim U, Lundman B. Qualitative content analysis in nursing research: concepts, procedures and measures to achieve trustworthiness. Nurse Educ Today. 2004;24:105–12.
15. Meyenburg B. Geschlechtsdysphorie im Jugendalter. Schwierige Behandlungsverläufe. Prax Kinderpsychol Kinderpsychiat. 2014;63:528–37.
16. Garrels L, Kockott G, Michael N, Preuss W, Renter K, Schmidt G, et al. Sex ratio of transsexuals in Germany: the development over three decades. Acta Psychiatr Scand. 2000;102:445–8.
17. Kaltiala-Heino R, Fröjd S. Correlation between bullying and clinical depression in adolescent patients. Adolescent Health Med Therapeutics. 2011;2:37–44.
18. Larson R, Richards MH. Daily companionship in late childhood and early adolescence. Changing developmental contexts. Child Dev. 1991;62:284–300.
19. Hall-Lande J, Eisenberg M, Christenson S, Neumark-Sztainer D. Social isolation, psychological health, and protective factors in adolescence. Adolescence. 2007;42:265–86.
20. Laursen B, Hartl A. Understanding loneliness during adolescence: developmental changes that increase the risk of perceived social isolation. J Adolescence. 2013;36:1261–8.
21. Rubin KH, Root AK, Bowker J. Parents, peers, and social withdrawal in childhood: a relationship perspective. In: Gazelle H, Rubin KH, editors. Social anxiety in childhood: bridging developmental and clinical perspectives. New Directions for Child and Adolescent Development, 127, 79–94. San Francisco: Jossey-Bass; 2010.
22. Shevlin M, Murphy S, Mallett J, Stringer M, Murphy J. Adolescent loneliness and psychiatric morbidity in Northern Ireland. Br J Clin Psychol. 2013;52:230–4.
23. Harris R, Qualter P, Robinson S. Loneliness trajectories from middle childhood to pre-adolescence: impact on perceived health and sleep disturbance. J Adolescence. 2013;36:1295–304.
24. Matheson S, Vijayan H, Dickson H, Shepherd A, Carr V, Laurens K. Systematic meta-analysis of childhood social withdrawal in schizophrenia, and comparison with data from at-risk children aged 9–14 years. J Psychiatr Res. 2013;47:1061–8.
25. Kaltiala-Heino R, Fröjd S, Autio V, Laukkanen E, Närhi P, Rantanen P. Transparent criteria for specialist level adolescent psychiatric care. Eur Child Adolesc Psychiatr. 2007;16:260–70.
26. Isojoki I, Fröjd S, Rantanen P, Laukkanen E, Närhi P, Kaltiala-Heino R. Priority criteria tool for elective specialist level adolescent psychiatric care predicts treatment received. Eur Child Adolesc Psychiatr. 2008;17:397–405.
27. deVries AL, Doreleijers TA, Steensma TD, Cohen-Kettenis PT. Psychiatric comorbidity in gender dysphoric adolescents. J Child Psychol Psychiatry. 2011;52:1195–202.
28. deVries A, Noens I, Cohen-Kettenis P, van Berckelaer-Onnes I, Doreleijers T. Autism spectrum disorders in gender dysphoric children and adolescents. J Autism Dev Disord. 2010;40:930–6.
29. Lord C, Rutter M, DiLavore PS, Risi S. Autism Diagnostic Observation Schedule (ADOS). Los Angeles, CA: Western Psychological Services; 1999.
30. Skuse D, Warrington R, Bishop D, Chowdhury U, Lau J, Mandy W, et al. The developmental, dimensional and diagnostic interview (3Di): a novel computerized assessment for autism spectrum disorders. J Am Acad Child Adolesc Psychiatry. 2004;43:548–58.
31. Rutter M, Le Couteur A, Lord C. Autism diagnostic interview-revised. Western Psychological Services: Los Angeles, CA; 2003.
32. Fombonne E. Epidemiology of autistic disorder and other pervasive developmental disorders. J Clin Psychiatr. 2005;66 Suppl 10:3–8.
33. Harter S. The construction of self. New York: Guilford Press; 1999.
34. Erikson E. Childhood and society. Harmonswordth, UK: Penguin; 1965.
35. Moshman D. Adolescent rationality and development. Cognition, morality, and identity. Third Edition. New York: Psychology Press; 2011.
36. Savin-Williams RC, Ream GL. Prevalence and stability of sexual orientation components during adolescence and young adulthood. Arch Sex Behav. 2007;36:385–94.
37. Muuss RE. Theories of adolescence (6th edition). New York: McGraw-Hill; 1965.
38. Gardner M, Steinberg L. Peer influence on risk taking, risk preference, and risky decision making in adolescence and adulthood: an experimental study. Dev Psychol. 2005;41:625–35.
39. Sebastian C, Burnett S, Blakemore S-J. Development of the self-concept during adolescence. Trends Cognit Sc. 2008;12:441–6.

Effects of gender affirming therapies in people with gender dysphoria: evaluation of the best available evidence

Romina Brignardello-Petersen, DDS, MSc, PhD
Wojtek Wiercioch, MSc, PhD

1. Introduction

We prepared this report to fulfill a request from the Florida Agency for Health Care Administration. This report contains three documents: 1. Main document (this document) summarizing the methodology used and the findings, 2. Methods document, which provides a detailed description of the systematic methodology used to find, prioritize, appraise, and synthesize the evidence, and 3. Results document, which describes the evidence available, the estimates of the effects of gender affirming therapies, and the certainty (also known as quality) of the evidence.

This document is organized in four parts. First, we describe the credentials and expertise of the health research methodologists conducting this evidence evaluation. Second, we summarize the methodology used. Third, we summarize the main findings. Finally, we briefly discuss strengths and limitations of our process and of the evidence.

2. Credentials and expertise

Two experts in health research methodology, who specialize in evidence synthesis to support decision making, prepared this report. Their relevant credentials and expertise are described below.

Dr. Romina Brignardello-Petersen: Assistant Professor at the Department of Health Research Methods, Evidence, and Impact, at McMaster University. Dr. Brignardello-Petersen obtained a DDS degree (University of Chile) in 2007, an MSc degree in Clinical Epidemiology and Health Care Research (University of Toronto) in 2012, and MSc in Biostatistics (University of Chile) in 2015, and a PhD in Clinical Epidemiology and Health Care Research (University of Toronto) in 2016. Dr. Brignardello-Petersen has worked in evidence synthesis projects since 2010, and her research has focused on the methodology for the development of Systematic Reviews and Clinical Practice Guidelines since 2012. Through January 2022, she has published 122 peer reviewed scientific articles (24 as a first author and 9 as a senior author). Dr. Brignardello-Petersen has acted as a research methodologist for several groups and organizations, including the World Health Organization, the Pan-American Health Organization, the American Society of Hematologists, the American College of Rheumatology, and the Society for Evidence Based Gender Medicine, among others. Her research program has been awarded over $2M CAD from the Canadian Institutes for Health Research. Dr. Brignardello-Petersen has no lived experience as a person or family member of a person with gender dysphoria, and her research interests are not in this area.

Dr. Wojtek Wiercioch: Postdoctoral Research Fellow at the Department of Health Research Methods, Evidence, and Impact, at McMaster University. Dr. Wiercioch obtained an MSc degree (2014, McMaster University) and a PhD degree (2020, McMaster University) in Health Research Methodology. Dr. Wiercioch has worked in evidence syntheses projects since 2011, and his research focuses on evidence synthesis, guideline development methodology, and the guideline development process. Through April

2022, he has published 86 peer-reviewed scientific articles. Dr. Wiercioch has acted as a guideline methodologist for several groups and organizations, including the World Health Organization, the American Society of Hematologists, the Endocrine Society (of America), and the American Association for Thoracic Surgeons, among others. Dr. Wiercioch has no lived experience as a person or family member of a person with gender dysphoria, and his research interests are not in this area.

3. Methods

We conducted an overview of systematic reviews. We used a reproducible approach to search, select, prioritize, appraise, and synthesize the available evidence, following high methodological standards. We describe full details of the methodology in an accompanying document.

In brief, we searched for systematic reviews published in English language in Epistemonikos, OVID Medline, and grey literature sources, through April 30, 2022. We selected systematic reviews which included studies on young individuals with a diagnosis of gender dysphoria, who received puberty blockers, cross-sex hormones, or surgeries; and in which authors reported data regarding outcomes important to patients: gender dysphoria, depression, anxiety, quality of life, suicidal ideation, suicide, adverse effects, and complications. Systematic reviews could have included any type of primary study design.

The two reviewers screened all titles and abstracts, followed by full text of potentially relevant systematic reviews. We then prioritized the most useful systematic review providing evidence for each of the outcomes, using pre-established criteria that considered date of publication, applicability, availability of outcome data, methodological quality of the systematic review, and usefulness of the data synthesis conducted in the systematic review (see methods document for details).

After abstracting data from the systematic reviews, we synthesized the best available evidence for each of the outcomes, and assessed the certainty (also known as quality) of the evidence using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach. We conducted GRADE assessments using the information provided by the systematic review authors (risk of bias of primary studies, characteristics of included studies, results reported by the studies). We present the all the information about outcomes in GRADE summary of findings tables.

In addition, to evaluate the robustness of our conclusions, we systematically searched for and evaluated primary studies answering the questions of interest published after the authors of the included systematic reviews conducted their searches.

4. Results

We included 61 systematic reviews, from which 3 addressed the effects of puberty blockers, 22 addressed the effects of cross-sex hormones, 30 addressed the effects of surgeries, and 6 addressed the effects of more than one of these interventions. After our prioritization exercise, we included information from 2 systematic reviews on puberty blockers, 4 on cross-sex hormones, and 8 on surgeries.

4.1 Puberty blockers

For most outcomes (except suicidality), there is no evidence about the effect of puberty blockers compared to not using puberty blockers. In other words, no studies compared the outcomes between a group of people with gender dysphoria using puberty blockers and another group of people with gender dysphoria not using them. Therefore, it is unknown whether people with gender dysphoria who use puberty blockers experience more improvement in gender dysphoria, depression, anxiety, and quality of life than those with gender dysphoria who do not use them. There is very low certainty about the effects of puberty blockers on suicidal ideation.

The studies included in the systematic review reported outcomes among a group of people with gender dysphoria after receiving puberty blockers. Low certainty evidence suggests that after treatment with puberty blockers, people with gender dysphoria experience a slight increase in gender dysphoria, and an improvement in depression, and anxiety. Low certainty evidence also suggests that a moderate percentage of patients experience adverse effects. The findings must be interpreted considering that these studies did not have a comparison group, and that it is unknown if people with gender dysphoria that do not use puberty blockers experience similar or different outcomes.

4.2 Cross sex hormones

For almost all outcomes (except breast cancer) there is no evidence about the effect of cross sex hormones compared to not using cross sex hormones. In other words, no studies compared the outcomes between a group of people with gender dysphoria using cross sex hormones and another group of people with gender dysphoria not using them. Therefore, it is unknown whether people with gender dysphoria who use cross-sex hormones experience more improvement in gender dysphoria, depression, anxiety, quality of life, and suicidality than those with gender dysphoria who do not use cross-sex hormones. There is low certainty evidence suggesting that cross-sex hormones may not increase the risk of breast cancer.

The studies included in the systematic reviews reported changes in the outcomes among a group of patients with gender dysphoria after the use of cross-sex hormones. Low certainty evidence suggests that after treatment with cross-sex hormones, people with gender dysphoria experience an improvement in gender dysphoria, depression, anxiety, and suicidality. There is very low certainty evidence about the changes in quality of life. There is moderate certainty evidence suggesting a low prevalence of venous thromboembolism after treatment with cross-sex hormones. The findings must be interpreted considering that these studies did not have a comparison group, and that it is unknown if people with gender dysphoria that do not use cross-sex hormones experience similar or different outcomes.

4.3 Surgeries

There were no systematic reviews and studies reporting on gender dysphoria, depression, anxiety, and suicidality. Therefore, the effects of surgeries on these outcomes (when compared to a group of patients with gender dysphoria who do not undergo surgery), or the changes in these outcomes (improvements or deterioration) among patients who undergo any gender-affirming surgery is unknown. Because of the lack of comparative studies, it is also unknown whether people with gender dysphoria who undergo surgeries experience more improvement in quality of life or less regret than those with gender dysphoria who do not undergo any surgeries. There is low certainty evidence suggesting that a low percentage of participants experience regret, and very low certainty evidence about changes in quality of life after surgery.

In assigned females at birth, low certainty evidence suggests that a high percentage of people are satisfied after chest surgery. There is very low certainty evidence, however, about satisfaction after bottom surgery, and about complications after both chest and bottom surgery. In assigned males at birth, low certainty evidence suggests a high percentage of people satisfied and a low percentage of people experiencing regret after vaginoplasty. There is very low certainty, however, about satisfaction with chest surgery and complications and reoperations after bottom surgery.

4.4 Evidence published after the systematic reviews selected

We found 10 relevant studies that were published after the systematic reviews were conducted. This evidence was not sufficient to importantly change the conclusions previously made.

5. Discussion

5.1 Summary of the evidence

In this report, we systematically summarized the best available evidence regarding the effects of puberty blockers, cross-sex hormones, and surgeries in young people with gender dysphoria. We did not find evidence about the effect of these interventions on outcomes important to patients when compared to not receiving the intervention. We found low and very low certainty evidence suggesting improvements in gender dysphoria, depression, anxiety, and quality of life, as well as low rates of adverse events, after treatment with puberty blockers and cross-sex hormones.

5.2 Completeness and applicability

There are several gaps in the evidence regarding the effects of puberty blockers, cross-sex hormones, and surgeries in patients with gender dysphoria. Although we found some evidence for all the outcomes of interest, the evidence is suboptimal: several limitations included the lack of studies with a comparison group, and the risk of bias and imprecision, resulting in low or very low certainty evidence for all outcomes.

The applicability of the evidence may also be limited. Although we only rated down for indirectness when it was considered a serious problem (i.e., in evidence about the effects of surgeries, which was collected from people who were importantly older than the target population in this report), there are also potential applicability issues to consider in the evidence regarding the effects of puberty blockers and cross-sex hormones. It is not clear to what extent the people included in the studies were similar enough to the people seeking these treatment options today. For example, some of the included studies were conducted in people who had a diagnosis of gender dysphoria confirmed with strict criteria, as well as a supportive environment. It is important to take into account to what extent this may compromise the applicability of the results to people who are not in the same situation.

5.3 Strengths and limitations of the process for developing this report

We followed a reproducible process for developing this report. We used the highest methodological standards and the approach to evidence synthesis we generally use when supporting organizations in the development of their guidelines. This approach is based on prioritizing the sources of evidence most likely to be informative (i.e., to identify and use the evidence with the highest certainty level).

To follow the principles for evidence-based decision making, which require using the best available evidence to inform decisions, we summarized the best available evidence. Because knowing the best

available evidence necessitates being aware of all the available evidence, we based this report on systematic reviews of the literature. We chose the most trustworthy and relevant systematic reviews among many published reviews.

One potential limitation of the process is that, due to feasibility concerns, we relied on the information reported by the systematic reviewers. Most of the systematic reviews we used, unfortunately, were judged at moderate or low methodological quality, which may raise concerns about the trustworthiness of the evidence presented in this report. We believe, however, that the results and conclusions of this report would not be importantly different had the systematic reviews been conducted following higher methodological standards. Because there are no randomized controlled trials, well-conducted comparative observational studies, or very large case series (which include a large sample of consecutive patients who are representative of the whole population) addressing the effects of puberty blockers, cross-sex hormones, and surgeries; the certainty of the evidence about the effects of these interventions is likely to continue being low or very low, even if a few more studies are included (as observed after searching for primary studies published after the reviews were conducted) or some data points were reported inaccurately in the systematic reviews.

Also due to feasibility concerns, the scope of this report was limited to outcomes that are important to patients. Although some may question the decision of not including surrogate outcomes for which there is evidence available (e.g. bone density, blood pressure), decision makers should rarely consider these outcomes and should instead focus on outcomes that do matter to people and stakeholders (e.g., fractures, cardiovascular events).

5.4 Implications

The evidence evaluating the effects of puberty blockers, cross-sex hormones, and surgeries in people with gender dysphoria has important limitations. Therefore, decisions regarding their use should carefully consider other relevant factors. At a patient level, these factors include patients' values and preferences (how patients trade off the potential benefit and harms - what outcomes are more important to them), and resources needed to provide the interventions (and the availability of such resources). At a population level, in addition to these factors, it would be important to consider resources needed to implement the interventions, feasibility, acceptability by relevant stakeholders, and equity.

It is important to note that when there is low or very low certainty evidence, it is rarely appropriate to make decisions that will be applied to the majority of the patients (equivalent to strong recommendations). This implies, at the patient level, that shared decision making is a key part of the decision-making process. At a policy level, extensive debate may be needed.

6. Conclusions

Due to the important limitations in the body of evidence, there is great uncertainty about the effects of puberty blockers, cross-sex hormones, and surgeries in young people with gender dysphoria. This evidence alone is not sufficient to support whether using or not using these treatments. We encourage decision makers to be explicit and transparent about which factors play an important role in their decision, and how they are weighed and traded off.

Methods

To ensure completeness and feasibility of the evidence review, we used an approach in which we prioritized the types of studies according to the design that was more likely to provide the best available evidence. First, we searched for systematic reviews of the literature. Second, we appraised all existing systematic reviews to select the most trustworthy (highest methodological quality, most up-to-date, most applicable) from which to draw conclusions. Third, we used the information presented in the systematic reviews to abstract information regarding the effects of the interventions of interest. Fourth, we assessed the certainty of the evidence (also known as quality of the evidence) abstracted from the selected systematic reviews. We planned to search for primary studies if systematic reviews were not found.

Information sources: We searched for existing systematic reviews in:

1. Epistemonikos (https://www.epistemonikos.org), an electronic database that focuses on systematic reviews. We used a comprehensive search strategy based on the population, using the terms “gender dysphoria”, “gender identity disorder” and “transgender”. We conducted this search on April 23, 2022.
2. OVID Medline. We used a search strategy based on the population and the interventions of interest, as well as an adaptation of a filter for systematic reviews from the Health Information Research Unit at McMaster University. We conducted this search on April 23, 2022.
3. Grey literature: we conducted a manual search in the websites of specific health agencies: National Institutes for Health and Care Excellence (NICE), Agency for Healthcare Research and Quality (AHRQ), Canada’s Drug and Health Technology Agency (CADTH), and the website from the Society for Evidence-Based Gender Medicine (SEGM). We conducted these searches between April 27-30, 2022.

We used no date limits for the searches, but we did limit to systematic reviews published in English. Search strategies are available in Appendix 1.

Eligibility criteria: We included systematic reviews, which we defined as:

1. Reviews in which the authors searched for studies to include in at least one electronic database, and in which there were eligibility criteria for including studies and a methodology for assessing and synthesizing the evidence, or
2. Reviews in which the authors searched for studies to include in at least one electronic database, and although there was no description of eligibility criteria or methodology, the presentation of the results strongly suggested that the authors used systematic methods (e.g. flow chart depicting study selection, tables with the same information from all included studies, synthesis of data at the outcome level).

We screened systematic reviews using the following criteria for inclusion:

- ***Type of participants:*** Young individuals (< 25 years old) with a diagnosis of gender dysphoria/gender identity disorder. We included reviews in which authors used any label and diagnostic criteria for this condition. We included reviews in which the participants in the reported studies were older if it was the only evidence available for a specific question. We excluded reviews with mixed populations (i.e. with and without gender dysphoria) in which people without gender dysphoria constituted more than 20% of the total sample.

- ***Type of Interventions*:** Puberty blockers, cross-sex hormones, gender affirming surgeries. We included any type of puberty blockers and cross-sex hormones, provided with any regimen. We included the following surgeries: phalloplasty, vaginoplasty, and chest surgery (mastectomy or breast implants/augmentation). We only included these when they were performed for the first time (i.e., not revision surgeries).

- ***Type of comparison:*** When the systematic reviews included comparative studies, the comparator of interest was no intervention. Participants could have received psychotherapy or counselling as a cointervention (in both groups).

- ***Type of outcomes:*** Gender dysphoria*,* mental health outcomes (depression and anxiety), quality of life, suicidal ideation, suicide*,* adverse effects (for puberty blockers and cross-sex hormones only), and satisfaction, complications, reoperation, and regret (for surgeries only). We included any length of follow-up. We excluded surrogate outcomes such as blood pressure, bone mineral density, kidney or liver function test values, etc.

- ***Type of studies included in the systematic reviews****:* Any clinical study (studies in which the researchers recruited and measured outcomes in humans) regardless of study design. This included randomized clinical trials, comparative observational studies, and case series. Because we could not quantify effect measures, incidence, or prevalence, we excluded case reports.

We excluded systematic reviews published only in abstract format, and those that we could not retrieve in full text (no access through the McMaster University library, or open access online).

Selection process: The two reviewers screened all titles and abstracts independently and in duplicate, followed by screening of full texts of potentially eligible systematic reviews independently and in duplicate, using the systematic review online application Covidence (https://www.covidence.org). We solved disagreements by consensus.

To select the most useful systematic reviews among all of those that met the eligibility criteria, we used the following prioritization criteria:

1. Date of publication: we prioritized systematic reviews published within the last 3 years (2020-2022)

2. Match between eligibility criteria of the review and the question of interest: we prioritized reviews in which the authors specifically included the population, intervention, comparison, and outcomes of interest for this evidence review
3. Outcome data available: we prioritized systematic reviews in which the authors report outcome data
4. Methodological quality: we used a modified version of the items in AMSTAR 2.[1] We modified the items to ensure assessment of methodological rather than reporting quality (Table 1). We rated each systematic review as having high, moderate, low, or critically low methodological quality, according to the guidance from the developers of the tool.[1] We reached consensus on critical items that determined this rating (Table 1). We prioritized selection of systematic reviews with highest methodological quality.

For surgical interventions, in addition, we prioritized systematic reviews that covered all gender affirming surgeries (instead of focusing on a specific type of surgery).

We selected a systematic review specifically for each of the outcomes of interest. In other words, we chose the best systematic review to inform each outcome. Each systematic review, however, could inform more than one outcome.

Table 1: Items used to rate the methodological quality of the eligible systematic reviews

AMSTAR Item	Modification to measure methodological quality
1. Did the research questions and inclusion criteria for the review include the components of PICO?	Does the review have a clear question and are the eligibility criteria for studies consistent with the question?
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	No modification needed
3. Did the review authors explain their selection of the study designs for inclusion in the review?	No modification needed
4. Did the review authors use a comprehensive literature search strategy?	Did the authors search in at least 2 electronic databases, using a reproducible search strategy?
5. Did the review authors perform study selection in duplicate?	No modification needed
6. Did the review authors perform data extraction in duplicate?	No modification needed
7. Did the review authors provide a list of excluded studies and justify the exclusions?	No modification needed
8. Did the review authors describe the included studies in adequate detail?	No modification needed
9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	No modification needed

10. Did the review authors report on the sources of funding for the studies included in the review?	Did the review authors consider conflicts of interest and how they may have affected the results of the primary studies?
11. If meta-analysis was performed, did the review authors use appropriate methods for statistical combination of results?	Was the synthesis of evidence done appropriately? (outcome level, appropriate meta analysis or narrative synthesis)
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	Did authors use subgroup or sensitivity analysis to assess the effect of risk of bias in meta-analytic results? Likely not applicable to most cases
13. Did the review authors account for RoB in primary studies when interpreting/discussing the results of the review?	Did the review authors incorporate an assessment of risk of bias at the outcome level when drawing conclusions?
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	Did the review authors incorporate an assessment of heterogeneity at the outcome level when drawing conclusions?
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	Did the authors address publication bias? (regardless of whether synthesis was using a meta-analysis or narrative)
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	Did the authors report conflicts of interest and did they manage any existing conflict of interest appropriately?

Shaded items were items considered critical.

Data abstraction: We abstracted outcome data from each of the systematic reviews. To ensure feasibility, we used the data as reported by the authors of the review and did not re-abstract data from the primary studies. One reviewer abstracted the data and a second reviewer checked the data for accuracy.

Data synthesis: Using the systematic reviews prioritized, we synthesized the evidence at the outcome level. Because of the higher likelihood of it resulting in higher certainty of evidence (details below) for each outcome, when there was comparative data (i.e. comparison of outcomes between an untreated and a treated group) and non-comparative data (i.e. changes from before to after treatment in one group, or only outcomes after treatment), we prioritized comparative data.

We prioritized numerical results (i.e. magnitudes of effect) and reported estimates and their 95% confidence intervals (CIs). When results were not reported in that way, we calculated the estimates and CIs when systematic review authors provided sufficient information. When necessary, we assumed moderate correlation coefficients for the changes between baseline and follow up (coefficient= 0.4). When this information was not available we reported narratively the effect estimates and ranges.

When a specific study reported the same outcome measured by more than one scale, we chose the scale presented first. We highlighted situations when the results obtained with other scales were importantly different.

When the same outcome was reported by more than one study but we could not pool the results, we created narrative syntheses.

Certainty of evidence: For each outcome, we assessed the certainty of the evidence (also known as quality of the evidence) using the Grading of Recommendations Assessment, Development, and Evaluation (GRADE) approach.[2] The certainty of evidence can be rated as high, moderate, low, or very low (Table 2). For effects of interventions, the certainty of the evidence started as high and could be rated down due to serious concerns about risk of bias, inconsistency, indirectness, imprecision, and publication bias. For inferences about the effect of using a treatment versus no treatment, when there was no comparison group, we assessed risk of bias as very serious and rated down the certainty of the evidence 2 levels by default. We used the same principles when assessing the certainty of the evidence in estimates of prevalence or rates, but did not judge risk of bias as resulting in very serious concerns due to lack of a comparison group. For all assessments, we used the information presented by the authors of the systematic review (e.g. assessments of risk of bias of the included studies, effect estimates from studies).

Table 2: GRADE levels of certainty of the evidence

Certainty level	Definition
High ⊕⊕⊕⊕	We are very confident that the true result (effect estimate/ prevalence/ mean, etc.) lies close to that of the estimate of the result
Moderate ⊕⊕⊕◯	We are moderately confident in the result: the true result is likely to be close to the estimate of the result, bur there is a possibility that it is substantially different
Low ⊕⊕◯◯	Our confidence in the result is limited: the true result may be substantially different from the estimate of the result
Very low ⊕◯◯◯	We have very little confidence in the result: the true result is likely to be substantially different from the estimate of the result

Presentation of results: We created GRADE Summary of Findings tables in which we describe the evidence available for each of the outcomes, and the certainty of the evidence. These tables contain the following information:

- Outcomes: measurement method (including scales, if applicable) and follow-up
- Estimates of effect: absolute and relative estimates of effect, and their corresponding 95% CIs.
- Number of studies and participants providing evidence for the outcome
- GRADE certainty of the evidence, with a link to detailed explanations (provided at the bottom of the table) of why the certainty of the evidence was rated at a specific level
- A narrative statement about what happens with the outcome, based on the estimate of effect and certainty of evidence.

Searching for new evidence not included in the systematic reviews: To assess if newer evidence not included in the included systematic reviews would change the conclusions importantly, we searched for and assessed primary studies answering the questions of interest that were published after the authors of such systematic reviews conducted their searches. We defined an important change in conclusions as a change in the certainty of the evidence (from low/ very low/ not available to high/ moderate).

We searched OVID Medline from January 1, 2019 through May 12, 2022, for studies published in English. We included studies if they enrolled young individuals (< 25 years old, with at least 20% of the people being this age) with a diagnosis of gender dysphoria/gender identity disorder, who received puberty blockers, cross-sex hormones, or surgeries; and measured any of the outcomes of interest.

For outcomes that should be evaluated in a comparative manner (e.g., depression, anxiety, etc.), because they are the only type of study design that would change the conclusions importantly, we selected comparative clinical studies (studies in which the researchers recruited and measured outcomes in humans, and compared a group of people who received the intervention with another one who did not receive the intervention). This included randomized clinical trials, and comparative observational studies. For outcomes that can only occur when the treatment is administered, we included non-comparative observational studies (case series). For these to change conclusions, they should have a sufficiently large sample size, and therefore we excluded case series in which the researchers reported information from <100 people.

Two reviewers screened the potentially relevant articles at title and abstract and full text screening stage. We abstracted relevant study characteristics and outcome data, and assessed risk of bias of comparative studies using the most relevant domains of the Risk of Bias for non-Randomized studies of Interventions (ROBINS-I) tool[3] (table 3). For non-comparative studies, we used a list of custom items that captured the most important potential risk of bias concerns of case series (table 4). We judged the risk of bias of each study as the highest risk of bias of any of the domains assessed (e.g., one domain judged at critical risk of bias resulted in the study judged at critical risk of bias). We summarized this information at the study and judged whether it would have changed the conclusions importantly if added to the body of evidence from the systematic reviews.

Table 3: Domains used to assess risk of bias of comparative studies

Domain	Low	Critical
Confounding	Adjusted for all relevant confounding factors	No adjustment
Classification of intervention	Intervention recorded prospectively or from medical records	Asked patients to recall whether they received the intervention
Deviation from intended interventions	No cointerventions or cointerventions balanced between the groups	Cointerventions unbalanced between the groups

Missing data	More than 90% of patients who started the study provided outcome data	Less thank 50% of patients who started the study provided outcome data
Measurement of outcome	All outcomes measured in the same way in both groups	Outcomes measured differently in both groups

Each domain could be judged at low, moderate, serious, or critical risk of bias. In addition, information could be insufficient to make a judgment. The table describes the criteria used to judge a domain in the extreme categories.

Table 4: Domains used to assess risk of bias of non- comparative studies

Domain	Low	High
Representativeness of the sample	Included all consecutive patients	Highly selected sample based on specific characteristics related with the prognosis after treatment
Classification of the intervention	Intervention recorded prospectively or from medical records	Asked patients to recall whether they received the intervention
Deviation from intended interventions	No cointerventions outside what would be observed in practice (or in a small proportion of patients)	Most patients received co interventions that could influence the outcomes
Missing data	More than 90% of patients who started the study provided outcome data	Less thank 50% of patients who started the study provided outcome data
Measurement of outcome	Outcomes measured prospectively or from medical records	Outcomes reported by the patients and/ or needed to recall what happened a long time ago

Each domain could be judged at low, moderate, or high risk of bias. In addition, information could be insufficient to make a judgment. The table describes the criteria used to judge a domain in the extreme categories.

References

1. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *Bmj* 2017;358:j4008. doi: 10.1136/bmj.j4008 [published Online First: 2017/09/25]
2. Blashem H, Helfand M, Schunemann HJ, et al. GRADE guidelines: 3. Rating the quality of the evidence. *Journal of clinical epidemiology* 2011;64:401-06.
3. Sterne JA, Hernan MA, Reeves BC, et al. ROBINS-I: a tool for assessing risk of bias in non-randomised studies of interventions. *BMJ (Clinical research ed)* 2016;355:i4919. doi: 10.1136/bmj.i4919 [published Online First: 2016/10/14]

Search Strategies

Questions Covered:

PICO questions:

1. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of treatment with **puberty blockers (gonadotrophin releasing hormone (GnRH) analogues)** compared to no puberty blockers?

2. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of treatment with **cross-sex hormones** compared to no cross-sex hormones?

3. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of **gender-affirming surgeries** compared to no surgery?

Search Strategies:

Note: Population, puberty blocker, cross-sex hormones search blocks adapted from NICE (2020) evidence reviews. Gender-affirming search block adapted from Wernick *et al.* 2019. Systematic reviews filter adapted from McMaster University Health Information Research Unit (HIRU).

Databases: Medline, Epistemonikos
Grey Literature: CADTH, AHRQ, SEGM, NICE

Medline

OVERVIEW	
Interface:	Ovid
Databases:	OVID Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
Study Types:	Systematic Reviews
Search Run:	April 23, 2022

Search Strategy: search terms [number of results]

Population

1	exp "Sexual and Gender Minorities"/	12385
2	Gender Dysphoria/	774
3	Gender Identity/	20481
4	Gender Role/	197
5	"Sexual and Gender Disorders"/	81
6	Transsexualism/	4236
7	Transgender Persons/	5303
8	Health Services for Transgender Persons/	186

9 exp Sex Reassignment Procedures/ 1208
10 gender identity disorder.mp. 492
11 non-binary.mp. 566
12 transgender.mp. 9989
13 (gender* adj3 (dysphori* or disorder* or distress or nonconform* or non-conform* or atypical or incongru* or identi* or disorder* or confus* or minorit* or queer* or variant or diverse or creative or explor* or question* or expan* or fluid)).tw. 16428
14 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition* or expression*)).tw. 13749
15 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).tw. 19665
16 (genderfluid or genderqueer or agender).mp. 130
17 ((correct or chosen) adj3 name).mp. 591
18 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).tw. 135313
19 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition* or expression*)).tw. 13749
20 (male-to-female or m2f or female-to-male or f2m).tw. 148579
21 or/1-20 342948

Cross-Sex Hormones

22 Hormones/ad, tu, th 4676
23 exp Progesterone/ad, tu, th 11265
24 exp Estrogens/ad, tu, th 29635
25 exp Gonadal Steroid Hormones/ad, tu, th 35375
26 (progesteron* or oestrogen* or estrogen*).tw. 223307
27 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).tw. 1488
28 exp Estradiol/ad, tu, th 11197
29 exp Testosterone/ad, tu, th 8710
30 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).tw. 86509
31 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).tw. 100252
32 or/22-31 345895

Puberty Blockers

33 Gonadotropin-Releasing Hormone/ 28809
34 (pubert* adj3 block*).ti,ab. 141
35 ((gonadotrophin or gonadotropin) and releasing).ti,ab. 20121
36 (GnRH adj2 analog*).ti,ab. 2878
37 GnRH*.ti,ab. 24390
38 "GnRH agonist*".ti,ab. 4749
39 Triptorelin Pamoate/ 1981
40 triptorelin.ti,ab. 821
41 arvekap.ti,ab. 1

42	("AY 25650" or AY25650).ti,ab.	1
43	("BIM 21003" or BIM21003).ti,ab.	0
44	("BN 52014" or BN52014).ti,ab.	0
45	("CL 118532" or CL118532).ti,ab.	0
46	Debio.ti,ab.	119
47	diphereline.ti,ab.	28
48	moapar.ti,ab.	0
49	pamorelin.ti,ab.	1
50	trelstar.ti,ab.	3
51	triptodur.ti,ab.	1
52	("WY 42422" or WY42422).ti,ab.	0
53	("WY 42462" or WY42462).ti,ab.	0
54	gonapeptyl.ti,ab.	0
55	decapeptyl.ti,ab.	225
56	salvacyl.ti,ab.	0
57	Buserelin/	2137
58	buserelin.ti,ab.	1395
59	onist.ti,ab.	0
60	("hoe 766" or hoe-766 or hoe766).ti,ab.	72
61	profact.ti,ab.	2
62	receptal.ti,ab.	31
63	suprecur.ti,ab.	5
64	suprefact.ti,ab.	25
65	tiloryth.ti,ab.	0
66	histrelin.ti,ab.	78
67	"LHRH-hydrogel implant".ti,ab.	1
68	("RL 0903" or RL0903).ti,ab.	1
69	("SPD 424" or SPD424).ti,ab.	1
70	goserelin.ti,ab.	1016
71	Goserelin/	1643
72	("ici 118630" or ici118630).ti,ab.	51
73	("ZD-9393" or ZD9393).ti,ab.	0
74	zoladex.ti,ab.	388
75	leuprorelin.ti,ab.	525
76	carcinil.ti,ab.	0
77	enanton*.ti,ab.	26
78	ginecrin.ti,ab.	0
79	leuplin.ti,ab.	15
80	Leuprolide/	3018
81	leuprolide.ti,ab.	2004
82	lucrin.ti,ab.	16
83	lupron.ti,ab.	183
84	provren.ti,ab.	0
85	procrin.ti,ab.	3
86	("tap 144" or tap144).ti,ab.	41
87	(a-43818 or a43818).ti,ab.	3
88	Trenantone.ti,ab.	2
89	staladex.ti,ab.	0

90 prostap.ti,ab. 6
91 Nafarelin/ 327
92 nafarelin.ti,ab. 263
93 ("76932-56-4" or "76932564").ti,ab. 0
94 ("76932-60-0" or "76932600").ti,ab. 0
95 ("86220-42-0" or "86220420").ti,ab. 0
96 ("rs 94991 298" or rs94991298).ti,ab. 0
97 synarel.ti,ab. 13
98 deslorelin.ti,ab. 306
99 gonadorelin.ti,ab. 237
100 ("33515-09-2" or "33515092").ti,ab. 0
101 ("51952-41-1" or "51952411").ti,ab. 0
102 ("52699-48-6" or "52699486").ti,ab. 0
103 cetrorelix.ti,ab. 520
104 cetrotide.ti,ab. 52
105 ("NS 75A" or NS75A).ti,ab. 0
106 ("NS 75B" or NS75B).ti,ab. 0
107 ("SB 075" or SB075).ti,ab. 1
108 ("SB 75" or SB75).ti,ab. 67
109 gonadoliberin.ti,ab. 151
110 kryptocur.ti,ab. 7
111 cetrorelix.ti,ab. 520
112 cetrotide.ti,ab. 52
113 antagon.ti,ab. 18
114 ganirelix.ti,ab. 160
115 ("ORG 37462" or ORG37462).ti,ab. 3
116 orgalutran.ti,ab. 26
117 ("RS 26306" or RS26306).ti,ab. 5
118 ("AY 24031" or AY24031).ti,ab. 0
119 factrel.ti,ab. 13
120 fertagyl.ti,ab. 12
121 lutrelef.ti,ab. 5
122 lutrepulse.ti,ab. 3
123 relefact.ti,ab. 10
124 fertiral.ti,ab. 0
125 (hoe471 or "hoe 471").ti,ab. 6
126 relisorm.ti,ab. 4
127 cystorelin.ti,ab. 19
128 dirigestran.ti,ab. 5
129 or/33-128 47108

Gender-affirming Surgeries

130 virilization/ 2309
131 (virilism or virili?ation or masculini?ation).mp. 5657
132 feminization/ 797
133 femini?ation.mp. 3420
134 (vaginoplasty or vaginoplasties).mp. 1022

135 exp Vagina/ or *Reconstructive Surgical Procedures/ 78841
136 (vaginoplasty or vaginoplasties).mp. 1022
137 (phalloplasty or phalloplasties).mp. 561
138 exp Penile Prosthesis/ 1636
139 "penile reconstruction".mp. 292
140 (vagina reconstruction or vaginal reconstruction).mp. 549
141 (genitoplasty or genitoplasties).mp. 263
142 transsexualism/su [Surgery] 1007
143 sex reassignment.mp. 1668
144 sex transformation.mp. 42
145 or/130-144 91560

Systematic Review Filter

147 meta-analysis/ 158633
148 (meta anal* or meta-anal* or metaanal*).ti,ab. 231876
149 ((systematic or evidence) adj2 (review* or overview*)).ti,ab. 279806
150 ((pool* or combined) adj2 (data or trials or studies or results)).ab. 65411
151 (search strategy or search criteria or systematic search or study selection or data extraction).ab. 70886
152 (search* adj4 literature).ab. 84593
153 or/146-152 521554

Combine Interventions and Population

154 32 or 129 or 145 459771
155 21 and 154 17838

Limit to Systematic Reviews in English Language

156 153 and 155 295
157 limit 156 to english language 288

Epistemonikos

OVERVIEW	
Interface:	https://www.epistemonikos.org/
Database:	Epistemonikos
Study Types:	Systematic Reviews
Search Run:	April 23, 2022

Search Strategy: search terms [number of results]

Population

(title:((title:(gender dysphoria) OR abstract:(gender dysphoria)) OR (title:(gender identity disorder) OR abstract:(gender identity disorder)) OR (title:(transgender) OR abstract:(transgender))) OR abstract:((title:(gender dysphoria) OR abstract:(gender dysphoria)) OR (title:(gender identity disorder) OR abstract:(gender identity disorder)) OR (title:(transgender) OR abstract:(transgender))))

Limit to Systematic Reviews

*Limited by publication type "systematic review" [425]

Canadian Agency for Drugs and Technologies in Health (CADTH)

OVERVIEW	
Interface:	https://www.cadth.ca/
Database:	CADTH
Study Types:	Systematic Reviews, Health Technology Reviews
Search Run:	April 27, 2022

Search Strategy: search terms [number of results]

"gender dysphoria" [10]
Limit to Health Technology Review [2]

"transgender" [9]
Limit to Health Technology Review [5]

"gender identity disorder" [1]

Agency for Healthcare Research and Quality (AHRQ)

OVERVIEW	
Interface:	https://search.ahrq.gov/
Database:	AHRQ
Study Types:	Evidence Based Practice (EPC) Centre Reports, Full Research Reports, Health Technology Assessments
Search Run:	April 29, 2022
Search Strategy: search terms [number of results]	
Search titles only: "gender identity disorder" "gender dysphoria" "transgender" [7]	

Society for Evidence-based Gender Medicine (SEGM)

OVERVIEW	
Interface:	https://segm.org/news
Database:	SEGM News
Study Types:	Systematic Reviews
Search Run:	April 30, 2022
Search Strategy: search terms [number of results]	
Find in page: "systematic" [5]	

National Institute for Health and Care Excellence (NICE)

OVERVIEW	
Interface:	https://www.nice.org.uk/
Database:	NICE
Study Types:	Systematic Reviews, Guidelines with Systematic Reviews
Search Run:	April 30, 2022
Search Strategy: search terms [number of results]	
gender dysphoria [1] *Limit to Guidance* [1] transgender [10] *Limit to Guidance* [7]	

gender identity disorder [9]
Limit to Guidance [8]

Search Strategies – Individual Studies

Questions Covered:

PICO questions:

1. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of treatment with **puberty blockers (gonadotrophin releasing hormone (GnRH) analogues)** compared to no puberty blockers?

2. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of treatment with **cross-sex hormones** compared to no cross-sex hormones?

3. For children, adolescents, and young adults (<21) with gender dysphoria, what are the effects of **gender-affirming surgeries** compared to no surgery?

Search Strategies:

Note: Population, puberty blocker, cross-sex hormones search blocks adapted from NICE (2020) evidence reviews. Gender-affirming search block adapted from Wernick *et al.* 2019.

Databases: Medline

Medline

OVERVIEW	
Interface:	Ovid
Databases:	OVID Medline Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily and Ovid MEDLINE(R) 1946 to Present
Study Types:	Any
Search Run:	May 12, 2022

Search Strategy: search terms [number of results]

Population

1	exp "Sexual and Gender Minorities"/	12631
2	Gender Dysphoria/	781
3	Gender Identity/	20586
4	Gender Role/	204
5	"Sexual and Gender Disorders"/	81
6	Transsexualism/	4259
7	Transgender Persons/	5371
8	Health Services for Transgender Persons/	187
9	exp Sex Reassignment Procedures/	1211
10	gender identity disorder.mp.	492

11 non-binary.mp. 574
12 transgender.mp. 10079
13 (gender* adj3 (dysphori* or disorder* or distress or nonconform* or non-conform* or atypical or incongru* or identi* or disorder* or confus* or minorit* or queer* or variant or diverse or creative or explor* or question* or expan* or fluid)).ti,ab. 16546
14 ((sex or gender*) adj3 (reassign* or chang* or transform* or transition*)).ti,ab. 9375
15 (transgend* or transex* or transsex* or transfem* or transwom* or transma* or transmen* or transperson* or transpeopl*).ti,ab. 19788
16 (genderfluid or genderqueer or agender).mp. 132
17 ((correct or chosen) adj3 name).mp. 591
18 (trans or crossgender* or cross-gender* or crossex* or cross-sex* or genderqueer*).ti,ab. 135744
19 (male-to-female or m2f or female-to-male or f2m).ti,ab. 149067
20 or/1-19 341083

Cross-sex Hormones

21 Hormones/ad, tu, th 4690
22 exp Progesterone/ad, tu, th 11270
23 exp Estrogens/ad, tu, th 29646
24 exp Gonadal Steroid Hormones/ad, tu, th 35401
25 (progesteron* or oestrogen* or estrogen*).ti,ab. 223689
26 ((cross-sex or crosssex or gender-affirm*) and (hormon* or steroid* or therap* or treatment* or prescri* or pharm* or medici* or drug* or intervention* or care)).ti,ab. 1507
27 exp Estradiol/ad, tu, th 11200
28 exp Testosterone/ad, tu, th 8722
29 (testosteron* or sustanon* or tostran or testogel or testim or restandol or andriol or testocaps* or nebido or testavan).ti,ab. 86670
30 (oestrad* or estrad* or evorel or ethinyloestrad* or ethinylestrad* or elleste or progynova or zumenon or bedol or femseven or nuvelle).ti,ab. 100411
31 or/21-30 346508

Puberty Blockers

32 Gonadotropin-Releasing Hormone/ 28845
33 (pubert* adj3 block*).ti,ab. 142
34 ((gonadotrophin or gonadotropin) and releasing).ti,ab. 20158
35 (GnRH adj2 analog*).ti,ab. 2879
36 GnRH*.ti,ab. 24437
37 "GnRH agonist*".ti,ab. 4763
38 Triptorelin Pamoate/ 1983
39 triptorelin.ti,ab. 822
40 arvekap.ti,ab. 1
41 ("AY 25650" or AY25650).ti,ab. 1
42 ("BIM 21003" or BIM21003).ti,ab. 0
43 ("BN 52014" or BN52014).ti,ab. 0
44 ("CL 118532" or CL118532).ti,ab. 0

45	Debio.ti,ab.	119
46	diphereline.ti,ab.	28
47	moapar.ti,ab.	0
48	pamorelin.ti,ab.	1
49	trelstar.ti,ab.	3
50	triptodur.ti,ab.	1
51	("WY 42422" or WY42422).ti,ab.	0
52	("WY 42462" or WY42462).ti,ab.	0
53	gonapeptyl.ti,ab.	0
54	decapeptyl.ti,ab.	225
55	salvacyl.ti,ab.	0
56	Buserelin/	2137
57	buserelin.ti,ab.	1396
58	onist.ti,ab.	0
59	("hoe 766" or hoe-766 or hoe766).ti,ab.	72
60	profact.ti,ab.	2
61	receptal.ti,ab.	31
62	suprecur.ti,ab.	5
63	suprefact.ti,ab.	25
64	tiloryth.ti,ab.	0
65	histrelin.ti,ab.	78
66	"LHRH-hydrogel implant".ti,ab.	1
67	("RL 0903" or RL0903).ti,ab.	1
68	("SPD 424" or SPD424).ti,ab.	1
69	goserelin.ti,ab.	1017
70	Goserelin/	1644
71	("ici 118630" or ici118630).ti,ab.	51
72	("ZD-9393" or ZD9393).ti,ab.	0
73	zoladex.ti,ab.	388
74	leuprorelin.ti,ab.	529
75	carcinil.ti,ab.	0
76	enanton*.ti,ab.	26
77	ginecrin.ti,ab.	0
78	leuplin.ti,ab.	15
79	Leuprolide/	3018
80	leuprolide.ti,ab.	2003
81	lucrin.ti,ab.	16
82	lupron.ti,ab.	183
83	provren.ti,ab.	0
84	procrin.ti,ab.	3
85	("tap 144" or tap144).ti,ab.	41
86	(a-43818 or a43818).ti,ab.	3
87	Trenantone.ti,ab.	2
88	staladex.ti,ab.	0
89	prostap.ti,ab.	6
90	Nafarelin/	327
91	nafarelin.ti,ab.	263
92	("76932-56-4" or "76932564").ti,ab.	0

93 ("76932-60-0" or "76932600").ti,ab. 0
94 ("86220-42-0" or "86220420").ti,ab. 0
95 ("rs 94991 298" or rs94991298).ti,ab. 0
96 synarel.ti,ab. 13
97 deslorelin.ti,ab. 310
98 gonadorelin.ti,ab. 238
99 ("33515-09-2" or "33515092").ti,ab. 0
100 ("51952-41-1" or "51952411").ti,ab. 0
101 ("52699-48-6" or "52699486").ti,ab. 0
102 cetrorelix.ti,ab. 520
103 cetrotide.ti,ab. 52
104 ("NS 75A" or NS75A).ti,ab. 0
105 ("NS 75B" or NS75B).ti,ab. 0
106 ("SB 075" or SB075).ti,ab. 1
107 ("SB 75" or SB75).ti,ab. 67
108 gonadoliberin.ti,ab. 152
109 kryptocur.ti,ab. 7
110 cetrorelix.ti,ab. 520
111 cetrotide.ti,ab. 52
112 antagon.ti,ab. 18
113 ganirelix.ti,ab. 161
114 ("ORG 37462" or ORG37462).ti,ab. 3
115 orgalutran.ti,ab. 26
116 ("RS 26306" or RS26306).ti,ab. 5
117 ("AY 24031" or AY24031).ti,ab. 0
118 factrel.ti,ab. 13
119 fertagyl.ti,ab. 12
120 lutrelef.ti,ab. 5
121 lutrepulse.ti,ab. 3
122 relefact.ti,ab. 10
123 fertiral.ti,ab. 0
124 (hoe471 or "hoe 471").ti,ab. 6
125 relisorm.ti,ab. 4
126 cystorelin.ti,ab. 19
127 dirigestran.ti,ab. 5
128 or/32-127 47179

Surgery

129 virilization/ 2309
130 (virilism or virili?ation or masculini?ation).mp. 5664
131 feminization/ 798
132 femini?ation.mp. 3425
133 (vaginoplasty or vaginoplasties).mp. 1032
134 (vaginoplasty or vaginoplasties).mp. 1032
135 (phalloplasty or phalloplasties).mp. 561
136 exp Penile Prosthesis/ 1642
137 "penile reconstruction".mp. 292

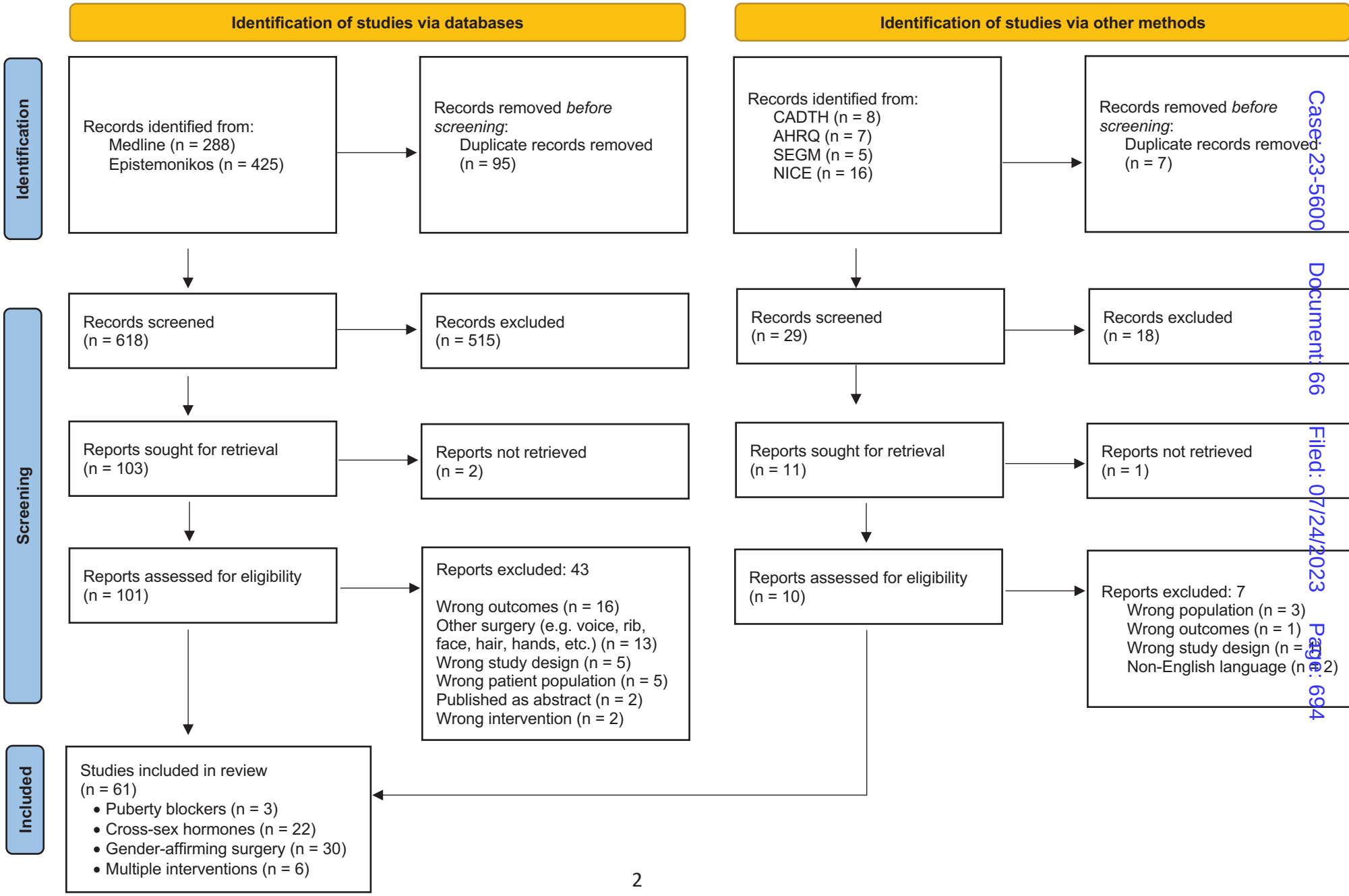
138	(vagina reconstruction or vaginal reconstruction).mp.	550
139	(genitoplasty or genitoplasties).mp.	263
140	transsexualism/su [Surgery]	1007
141	sex reassignment.mp.	1674
142	sex transformation.mp.	42
143	or/129-142	14290
	Any intervention AND population	
144	31 or 128 or 143	386835
145	20 and 144	16516
	Limit to Humans	
146	animals/ not humans/	4972586
147	145 not 146	9281
148	limit 147 to humans	7901
	Limit to Publication Year 2019 to Current	
149	limit 148 to yr="2019 -Current"	1859

Results

Search results and eligible reviews: After screening 647 records found through our searches, we found 61 eligible systematic reviews. From these, 27 were published between 2020 and 2022 (Figure 1). Overall, 4% (1/27) of the reviews were judged to be of high methodological quality, 15% (4/27) were moderate methodological quality, 37% (10/27) were low methodological quality, and 44% (12/27) were critically low methodological quality.

We provide reasons for excluding systematic reviews in appendix 1.

Effects of gender affirming therapies in people with gender dysphoria: evaluation of the best available evidence. Dr. Romina Brignardello-Petersen and Dr. Wojtek Wiercioch; Results; May 16, 2022



Effects of gender affirming therapies in people with gender dysphoria: evaluation of the best available evidence. Dr. Romina Brignardello-Petersen and Dr. Wojtek Wiercioch; Results; May 16, 2022

Figure 1: PRISMA flow diagram for the selection of systematic reviews. *From:* Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: http://www.prisma-statement.org/

Outcomes:

1. **Puberty blockers:** We found 4 systematic reviews assessing the effects of puberty blockers published between 2020 and 2022.[1-4] From these, we judged 2 as having moderate methodological quality, and 2 as having critically low methodological quality. Details of the assessment are provided in Figure 2.

   Table 1 summarizes the evidence about the effects of puberty blockers on the outcomes of interest. We used information from 2 systematic reviews.[2 3] For most outcomes (except suicidality), there is no evidence about the effect of puberty blockers compared to not using puberty blockers. In other words, no studies compared the outcomes between a group of people with gender dysphoria using puberty blockers and another not using them. Therefore, it is unknown whether people with gender dysphoria who use puberty blockers experience more improvement in gender dysphoria, depression, anxiety, and quality of life than those with gender dysphoria who do not use them. There is very low certainty about the effects of puberty blockers on suicidal ideation (see details in Table 1).

   Studies, however, reported outcomes among a group of people with gender dysphoria after receiving puberty blockers. The findings are:
   - There is low certainty evidence suggesting that treatment with puberty hormones may slightly increase gender dysphoria severity (mean change score in the Utrecht Gender Dysphoria scale, 0.7 points [95% CI, -4.2 to 5.6], range 12-60, with higher scores reflecting more severe gender dysphoria)
   - There is low certainty evidence suggesting that treatment with puberty blockers may decrease depression (mean change score in the Beck Depression Inventory, -3.4 [95% CI, -5.7 to -1.0], range 0-63, with higher scores reflecting more severe depression)
   - There is low certainty evidence suggesting that treatment with puberty blockers may decrease anxiety (mean change score in the Trait Anxiety Scale, trait subscale, -1.5 [95% CI, -4.7 to -1.8], range 0-80, with higher scores reflecting more severe anxiety)
   - There is low certainty evidence suggesting a moderate percentage of patients reporting adverse events after treatment with puberty blockers (see Table 1 for details)
   - There is very low certainty evidence about how puberty blockers affect suicidality

Figure 2: AMSTAR assessment judgements for systematic reviews addressing puberty blockers

Review ID	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Methodological quality
AHRQ 2021																	MODERATE
NICE 2020a																	MODERATE
Ramos 2020																	CRITICALLY LOW
Rew 2020																	CRITICALLY LOW

Figure legend:

Yes

Probably yes

Probably no

No

Not applicable

Table 1: Puberty blockers (gonadotrophin releasing hormone analogues) compared to no puberty blockers in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: puberty blockers (gonadotrophin releasing hormone analogues)
Comparison: no puberty blockers

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no puberty blockers	**Risk / mean with puberty blockers**				
Gender dysphoria assessed with: difference (effect) in gender dysphoria proportion or severity	Not reported					The effects of puberty blockers on gender dysphoria are unknown
Gender dysphoria assessed with: mean change score in the Utrecht Gender Dysphoria Scale (12-60, higher scores reflect more gender dysphoria, 40 points or more indicate a diagnosis of gender dysphoria) (NICE, 2020a) Follow up: mean 1.9 years (range 0.4 to 5.1 years)	NA	0.7 (-4.2 to 5.6)	NA	41 (1 study)	⨁⨁◯◯ LOW[1]	The mean gender dysphoria score may increase by 0.7 points after puberty blockers
Depression assessed with: difference (effect) in depression proportion or severity	Not reported					The effects of puberty blockers on depression are unknown

Table 1: Puberty blockers (gonadotrophin releasing hormone analogues) compared to no puberty blockers in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: puberty blockers (gonadotrophin releasing hormone analogues)
Comparison: no puberty blockers

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no puberty blockers	**Risk / mean with puberty blockers**				
Depression assessed with: mean change score in Beck Depression Inventory-II scale (0-63, higher scores represent more severe depression) (NICE, 2020a) Follow up: mean 1.9 years (range 0.4 to 5.1 years)	NA	-3.4 (-5.7 to -1.0)	NA	41 (1 study)	⨁⨁◯◯ LOW[1]	The mean depression score may decrease by 3.4 points after puberty blockers
Anxiety assessed with: difference (effect) in anxiety proportion or severity	Not reported					The effects of puberty blockers on anxiety are unknown
Anxiety assessed with: mean change score in STAI-Trait scale (0-80, higher scores represent more severe anxiety) (NICE, 2020a) Follow up: mean 1.9 years (range 0.4 to 5.1 years)	NA	-1.5 (-4.7 to 1.8)	NA	41 (1 study)	⨁⨁◯◯ LOW[1]	The mean anxiety score may decrease by 1.5 points after puberty blockers

Table 1: Puberty blockers (gonadotrophin releasing hormone analogues) compared to no puberty blockers in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: puberty blockers (gonadotrophin releasing hormone analogues)
Comparison: no puberty blockers

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no puberty blockers	**Risk / mean with puberty blockers**				
Quality of life assessed with: any measure	Not reported					The effects of puberty blockers on quality of life are unknown
Suicidal ideation difference (effect) in suicidal ideation (Rew, 2020) Follow-up: cross-sectional survey	The authors report that "compared to youth who did not receive pubertal suppression, those who did showed lower lifetime rates of suicidal ideation".			89 (1 study)	⨁◯◯◯ VERY LOW[2]	We are very uncertain about the effect of puberty blockers on suicidal ideation
Adverse effects assessed with: proportion of patients reporting adverse effects (NICE, 2020a) Follow up: mean 2.3 years (range 0.0 to 11.3 years)	NA	11%[3] (2% to 29%)	NA	27 (1 study)	⨁⨁◯◯ LOW[4]	The proportion of patients reporting adverse effects after treatment with puberty blockers may be 11%

STAI-Trait: Trait Anxiety Scale. Range: 0-80
CI: Confidence interval
NA: Not applicable

Table 1: Puberty blockers (gonadotrophin releasing hormone analogues) compared to no puberty blockers in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: puberty blockers (gonadotrophin releasing hormone analogues)
Comparison: no puberty blockers

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no puberty blockers	**Risk / mean with puberty blockers**				

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

1. Mean change rated down due to risk of bias and imprecision. According to the systematic review authors, the study had poor methodological quality. In addition, there are too few participants included, which is not sufficient to make trustworthy inferences (does not meet the optimal information size).
2. The authors of Rew 2020 narratively summarized the outcome of Turban *et al.* 2020; a cross-sectional online survey study. According to the systematic review authors, Turban et al. did not describe the study participants and the setting in detail and it was unclear whether outcomes were measured in a valid and reliable way. We therefore, downgraded the certainty of evidence by one level from low to very low due to high risk of bias.
3. The authors reported 3/27 (11%) participants treated with GnRHa developed side effects: 1 participant developed sterile abscesses; they were switched from leuprolide acetate to triptorelin, 1 participant developed leg pains and headaches, which eventually resolved without treatment, 1 participant gained 19 kg within 9 months of initiating GnRH analogues.
4. Proportion of adverse effects rated down due to risk of bias and imprecision. According to the systematic review authors, the cohort study Khatchadourian *et al.* 2014 was assessed at high risk of bias due to incomplete reporting of its cohort. In addition, there are too few participants included, which is not sufficient to make trustworthy inferences (does not meet the optimal information size).

2. **Cross-sex hormones:** We found 9 systematic reviews assessing the effects of cross-sex hormones published between 2020 and 2022.[4-12] One of these, however, included both puberty blockers and cross-sex hormones combined in their evidence synthesis as was not prioritized.[5] From the 8 remaining reviews, we judged 1 as having high methodological quality, 2 as having moderate methodological quality, 2 as having low methodological quality, and 3 as having critically low methodological quality. Details of the assessment are provided in Figure 3. Because of its eligibility criteria related to study design, the systematic review judged at high methodological quality[7] did not include any studies and therefore we could not use it to inform any outcome.

   Table 2 summarizes the evidence about the effects of cross-sex hormones on the outcomes of interest. We used information from 4 systematic reviews.[6 9 11 12] For most outcomes (all except risk of breast cancer), there is no evidence about the effect of cross-sex hormones compared to not using cross-sex hormones. In other words, no studies compared the outcomes between a group of people with gender dysphoria using cross-sex hormones and another not using it. Therefore, it is unknown whether people with gender dysphoria who use cross-sex hormones experience more improvement in gender dysphoria, depression, anxiety, quality of life, and suicidality than those with gender dysphoria who do not use them. There is low certainty evidence suggesting that cross-sex hormones may not increase or decrease the risk of breast cancer (see details in Table 2).

   Studies, however, reported outcomes among a group of people with gender dysphoria after receiving cross-sex hormones. The findings are:
   - There is low certainty evidence suggesting that treatment with cross-sex hormones may decrease gender dysphoria severity (mean change score in the Utrecht Gender Dysphoria scale, -42.4 points [95% CI, -44.1 to -40.1], range 12-60, with higher scores reflecting more severe gender dysphoria)
   - There is low certainty evidence suggesting that treatment with cross-sex hormones may decrease depression (measured with different scales, see Table 4 for details) and the need for treatment for depression (change in percentage, -39%)
   - There is low certainty evidence suggesting that treatment with cross-sex hormones may decrease anxiety (measured with different scales, see Table 4 for details) and the need for treatment for anxiety (change in percentage, -32%)
   - There is very low certainty about the change in quality of life after treatment with cross-sex hormones.
   - There is low certainty evidence suggesting that treatment with cross-sex hormones may decrease suicidality degree (mean change score in the Ask Suicide-Screening questions scale, -0.84 points [95% CI, -1.30 to -0.44], range 0-4, with higher scores reflecting more severe suicidality) and the percentage of patients with need for treatment due to suicidality/self-harm (change in percentage, -31%). There is very low certainty evidence about the percentage of people with suicidal ideation and suicide attempts after treatment with cross-sex hormones.

- There is low certainty evidence suggesting a low prevalence of venous thromboembolism after treatment with cross-sex hormones (see Table 2 for details)

Figure 3: AMSTAR assessment judgements for systematic reviews addressing cross-sex hormones

Review ID	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Methodological quality
AHRQ 2021																	MODERATE
Baker 2021																	MODERATE
Fledderus 2020																	CRITICALLY LOW
Haupt 2020																	HIGH
Karalexi 2020																	LOW
Kotamarti 2021																	CRITICALLY LOW
Mattawanon 2021																	CRITICALLY LOW
NICE 2021b																	MODERATE
Totaro 2021																	LOW

Figure legend:

Yes

Probably yes

Probably no

No

Not applicable

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	**Risk/ mean with cross-sex hormones**				
Gender dysphoria assessed with: difference (effect) in gender dysphoria percentage or severity	Not reported					The effects of cross-sex hormones on gender dysphoria are unknown
Gender dysphoria assessed with: mean change score in the Utrecht Gender Dysphoria Scale (12-60, higher scores reflect more gender dysphoria, 40 points or more indicate a diagnosis of gender dysphoria) (NICE, 2020b) Follow up: 1 year	NA	-42.4 (-44.1 to -40.1)	NA	23 (1 study)	⨁⨁◯◯ LOW[1]	The mean gender dysphoria score may decrease by 42 points after cross-sex hormones
Depression assessed with: difference (effect) in depression percentage or severity	Not reported					The effects of cross-sex hormones on depression are unknown

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	**Risk/ mean with cross-sex hormones**				
Depression assessed with: mean change score in depression scales (higher scores represent more severe depression) (NICE, 2020b) Follow up: 1 year	NA	The mean depression score reduction was 9.6 points when using the BDI-II scale (n=23) and 7.5 when using the CESD-R scale (n=50). The authors report that both reductions were statistically significant[2]	NA	73 (2 studies)	⨁⨁◯◯ LOW[1]	The mean depression score may decrease after cross-sex hormones
Depression assessed with: change in percentage of patients with need for treatment (NICE, 2020b) Follow-up: 1 year	NA	The percentage of participants requiring treatment was reduced by 39% (from 54% at baseline), which was statistically significant	NA	52 (1 study)	⨁⨁◯◯ LOW[1]	The percentage of participants requiring treatment may be reduced by 39% after cross-sex hormones
Anxiety assessed with: difference (effect) in anxiety percentage or severity	Not reported					The effects of cross-sex hormones on anxiety are unknown

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				
Anxiety assessed with: mean change score in anxiety scales (higher scores represent more severe anxiety) (NICE, 2020b) Follow up: 1 year	NA	The mean anxiety score reduction was 16.5 points when using the STAI-State scale and 14.5 when using the STAI-Trait scale. The authors report that both reductions were statistically significant	NA	23 (1 study)	⨁⨁◯◯ LOW[1]	The mean anxiety score may decrease after cross-sex hormones
Anxiety assessed with: change in percentage of patients with need for treatment (NICE, 2020b) Follow-up: 1 year	NA	The percentage of participants requiring treatment was reduced by 32% (from 48% at baseline), which was statistically significant	NA	52 (1 study)	⨁⨁◯◯ LOW[1]	The percentage of participants requiring treatment may be reduced by 32% after cross-sex hormones
Quality of life assessed with: difference (effect) in quality of life improvement	Not reported					The effects of cross-sex hormones on quality of life are unknown

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	**Risk/ mean with cross-sex hormones**				
Quality of life assessed with: mean change score in QLES-Q-SF score (higher scores represent better quality of life) (NICE, 2020b) Follow up: 1 year	NA	The mean quality of life score improved, but the differences were not statistically significant. The magnitudes were not reported	NA	50 (1 study)	⨁◯◯◯ VERY LOW[3]	We are very uncertain about the quality of life change after cross-sex hormones
Suicide/ suicidal ideation assessed with: difference (effect) in suicide or suicidal ideation	Not reported					The effects of cross-sex hormones on suicide/ suicidal ideation are unknown
Suicidality assessed with: change in score from ASQ instrument (higher scores represent greater degree of suicidality) (NICE, 2020b) Mean follow up: 1 year	NA	-0.84 (-1.30 to -0.44)	NA	39 (1 study)	⨁⨁◯◯ LOW[1]	Suicidality scores may decrease by 0.84 points after cross-sex hormones

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				
Suicidal ideation assessed with: percentage of participants with suicidal ideation measured with PHQ-9 (NICE, 2020b) Follow-up: 1 year	NA	The percentage of participants with suicidal ideation decreased by 6% (from 10% at baseline). The authors did not conduct a statistical analysis	NA	50 (1 study)	⨁◯◯◯ VERY LOW[3]	We are very uncertain about the change in percentage of patients in suicidal ideation after cross-sex hormones
Suicide attempts assessed with: not reported (NICE, 2020b) Follow up: not reported	NA	The percentage of people with lifetime suicide attempts was 15%, those with attempts 3 months before treatment was 2%, and those with attempts at follow up was 5%	NA	130 (1 study)	⨁◯◯◯ VERY LOW[3]	We are very uncertain about the percentage of people with suicide attempts after cross-sex hormones
Suicidality/ self-harm assessed with: change in percentage of patients with need for treatment (NICE, 2020b) Follow-up: 1 year	NA	The percentage of participants requiring treatment was reduced by 31% (from 35% at baseline), which was statistically significant	NA	52 (1 study)	⨁⨁◯◯ LOW[1]	The percentage of participants requiring treatment may be reduced by 31% after cross-sex hormones
Venous thromboembolism assessed with: Risk of VTE	Not reported					The effects of cross-sex hormones on the risk of VTE are unknown

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				
Venous thromboembolism assessed with: Prevalence among assigned males at birth (Totaro, 2021) Mean follow up: 4.1 years	NA	**20 per 1,000** (10 to 30)	NA	11,542 (18 studies)	⨁⨁⨁◯ MODERATE[4]	The prevalence of VTE among assigned males at birth is probably 2% after cross-sex hormones
Venous thromboembolism assessed with: Prevalence among assigned females at birth (Kotamarti, 2021) Mean follow up: 5.7 years	NA	**6 per 1,000** (CI not reported)[5]	NA	4,218 (8 studies)	⨁⨁⨁◯ MODERATE[6]	The prevalence of VTE among assigned females at birth is probably 0.6% after cross-sex hormones
Breast cancer assessed with: Risk of breast cancer (Fledderus, 2020) Follow up: not reported	Two studies compare the risk of beast cancer between assigned females at birth using versus not using testosterone, and found no differences (0 vs 1 case [total n= 130], and 1 vs 6 [total n=1579]). A third study compared assigned females at birth with non transgender women and found a lower risk in the former (magnitude not reported)		NA	2,938 (3 studies)	⨁⨁◯◯ LOW[7]	The risk of breast cancer may not increase or decrease due to the use of cross-sex hormones

Table 2: Cross-sex hormones compared to no cross-sex hormones in youth (<21 years old) with gender dysphoria

Patient or population: youth (<21 years old) with gender dysphoria
Intervention: cross-sex hormones
Comparison: no cross-sex hormones

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no cross-sex hormones	Risk/ mean with cross-sex hormones				

ASQ: Ask Suicide-Screening Questions. Range 0-4
BDI-II: Beck Depression Inventory. Range: 0-63
CESD-R: Center for Epidemiological Studies Depression Scale. Range: 0-60
CI: Confidence interval
NA: Not applicable
PHQ-9: Patient Health Questionnaire (PHQ) Modified for Teens. For suicidal ideation, it is a single question (yes/no)
QLES-Q-SF: Quality of Life Enjoyment and Satisfaction Questionnaire. Range: 15-75
STAI: State-Trait Anxiety Inventory. Range: 0-80

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

1. Mean change rated down due to risk of bias and imprecision. According to the systematic review authors, the studies had poor methodological quality. In addition, there are too few participants included, which is not sufficient to make trustworthy inferences (does not meet the optimal information size)
2. Similar results when this outcome was measured using the Patient Health Questionnaire (PHQ) Modified for Teens in one of the same studies
3. Rated down due to risk of bias, imprecision, and indirectness. According to the systematic review authors, the studies had poor methodological quality. In addition, there are too few participants included, which is not sufficient to make trustworthy inferences (does not meet the optimal information size). Finally, 30% of the participants did not have a diagnosis of gender dysphoria.
4. Prevalence rated down due to risk of bias. According to the systematic review authors, only 6 out of the 18 studies (representing 16.5% of the weight of the studies) were at low risk of bias.

5. A meta-analysis of independent studies reported in this systematic review suggested that the prevalence of VTE in non-transgender females at birth was 1.7% (based on 7 studies and 18,748 persons)
6. Prevalence rated down doe to risk of bias. According to the systematic review authors, all studies had at least one domain judged as problematic.
7. Risk rated down 2 levels because of risk of bias. The researchers did not account for confounding in any of the studies.

3. **Surgeries:** We found 15 systematic reviews assessing the effects of gender-affirming surgeries published between 2020 and 2022. We judged 8 as having low methodological quality and 7 as having critically low methodological quality. Details of the assessment are provided in Figure 4. We present the results regarding the effects of surgeries in three parts. First, we describe the effects of all surgeries on mental health outcomes in all patients. Second, we describe the effects of all surgeries on surgical outcomes in assigned females at birth (transgender males). Finally, we describe the effects of all surgeries on surgical outcomes in assigned males at birth (transgender females).

3.1 Effects of surgeries on mental health outcomes: Table 3 summarizes the evidence about the effects of all surgeries on mental health outcomes in all patients. We used information from 2 systematic reviews.[13 14] There were no systematic reviews and studies reporting on gender dysphoria, depression, anxiety, and suicidality. Therefore, the effects of surgeries on these outcomes (when compared to a group of patients with gender dysphoria who do not undergo surgery), or the changes in these outcomes (improvements or deterioration) among patients who undergo surgeries is unknown.

The systematic reviews addressed quality of life and depression, but none of the included studies included a comparison group. Thus, it is unknown whether people with gender dysphoria who undergo surgeries experience more improvement in quality of life or less regret than those with gender dysphoria who do not undergo surgeries.

Studies, however, reported the following outcomes among a group of people with gender dysphoria after undergoing surgeries. The findings are:

- There is low certainty evidence suggesting that the percentage of people who experience regret after surgery is low (1%)
- There is very low certainty evidence about how surgeries affect quality of life (see Table 3 for details)

Figure 4: AMSTAR assessment judgements for systematic reviews addressing gender-affirming surgery

Review ID	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Methodological quality
Bustos SS 2021																	LOW
Bustos VP 2021																	LOW
Bustos VP 2021b																	LOW
Dunford 2021																	LOW
Eftekhar, 2020																	LOW
Falcone 2021																	CRITICALLY LOW
Hu, 2022																	CRITICALLY LOW
Huayllani 2021																	CRITICALLY LOW
Jolly 2021																	LOW
Nassiri 2020																	CRITICALLY LOW
Oles 2022																	LOW
Oles 2022b																	LOW
Salibian 2021																	CRITICALLY LOW
Sijben 2021																	CRITICALLY LOW
Tay 2021																	CRITICALLY LOW

Figure legend:

Yes

Probably yes

Probably no

No

Not applicable

Table 3: All surgeries compared to no surgeries in young people (<21 years old) with gender dysphoria

Patient or population: young people (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries
Outcomes: Mental health and regret

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Gender dysphoria assessed with: any measure	Not reported					The effects of surgery on gender dysphoria, the changes in gender dysphoria severity after surgery, and the prevalence of gender dysphoria after surgery are unknown
Depression assessed with: any measure	Not reported					The effects of surgery on depression, the changes in depression severity after surgery, and the prevalence of depression after surgery are unknown
Anxiety assessed with: any measure	Not reported					The effects of surgery on anxiety, the changes in anxiety severity after surgery, and the prevalence of anxiety after surgery are unknown
Suicidality assessed with: any measure	Not reported					The effects of surgery on suicidality, the changes in anxiety severity after surgery, and the prevalence of anxiety after surgery are unknown
Quality of life assessed with: difference (effect) in quality of life	Not reported					The effects of surgery on quality of life are unknown
Quality of life assessed with: change in quality of life	Not reported					The change in quality of life after surgery is unknown

Quality of life assessed with: mean score in the Short Form-36 Scale (0-100, higher scores reflect better quality of life) (Eftekhar Ardebili, 2020) Follow up: cross-sectional	NA	59.17 (48.59 to 69.74)[1]	NA	633 (5 studies)	⨁◯◯◯ VERY LOW[2]	We are very uncertain about the quality of life after surgeries
Regret assessed with: difference (effect) in percentage of people with regret	Not reported					The effects of surgery on regret are unknown
Regret assessed with: percentage of people with regret (Bustos, 2021) Mean follow up: 4 years	NA	1% (0 to 2%)[3]	NA	7928 (27 studies)	⨁⨁◯◯ LOW [4]	The percentage of people who experience regret is low

CI: Confidence interval
NA: Not applicable

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

1. Similar scores for assigned males at birth and assigned females at birth.
2. Mean score rated down for risk of bias and inconsistency. According to the systematic review authors, all studies had concerns related to risk of bias. In addition, the smaller studies showed better quality of life than the larger study.
3. Similar percentage for assigned males at birth and assigned females at birth, and for different types of surgeries (all pooled percentages below 2%).
4. Percentage rated down due to risk of bias and indirectness. According to the authors, many of the studies had moderate or high risk of bias. The mean age of the participants at the time of surgery was higher than the target population. Because it was considered to not have an important effect on the pooled estimate, we did not rate down for statistical heterogeneity

3.2 **Effects of surgeries on assigned females at birth:** Table 4 summarizes the evidence about the effects of all surgeries on surgical outcomes among assigned at birth females. We used information from 3 systematic reviews.[13-17] Due to the nature of the outcomes (i.e. they can only be experienced by people who undergo surgeries), there cannot be studies comparing the outcomes between a group of people with gender dysphoria who undergo surgeries and another who does not.

Studies, therefore, assessed the outcomes among a group of people with gender dysphoria after surgery. The findings are:

- There is low certainty evidence suggesting that the percentage of people who are satisfied after chest surgery is high (92%)
- There is very low certainty evidence about the rate of surgical complications after chest surgery
- There is very low certainty evidence about the percentage of people who are satisfied, and the rate of surgical complications after bottom surgeries (see Table 4 for details)

Table 4: All surgeries compared to no surgeries in assigned females at birth (<21 years old) with gender dysphoria

Patient or population: assigned females at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Chest surgery						
Satisfaction assessed with: percentage of people who reported being satisfied (Bustos VP, 2020b) Range of follow up: 6 weeks to 46 months[1]	NA	92% (88% to 96%)[2]	NA	733 (14 studies)	⨁⨁◯◯ LOW[3]	The percentage of people who reports being satisfied may be 92%
Surgical complications assessed with: rate of complications across patients (Oles, 2022) Range of follow up: 8 weeks to 1 year	NA	16.8% Range (5.5% to 80.0%)	NA	1255 (7 studies)	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the rate of surgical complications
Reoperation assessed with: rate of reoperation across patients (Oles, 2022) Range of follow up: 8 weeks to 1 year	NA	6.2% Range (0.7% to 11.2%)	NA	1214 (6 studies)	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the rate of reoperation
Bottom surgery						

Table 4: All surgeries compared to no surgeries in assigned females at birth (<21 years old) with gender dysphoria

Patient or population: assigned females at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Satisfaction assessed with: percentage of people who reported being satisfied (Oles, 2022b) Range of follow up: 6 weeks to 46 months	NA	89.6% (45% to 100%)[5]	NA	1458 (27 studies)	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the percentage of people who reports being satisfied
Surgical complications-Major assessed with: percentage of people experiencing major complications (Oles, 2022b) follow up: not reported	NA	The percentage was - 2.3% (range 0 to 20%) experiencing total flap loss - 19.5% (range 0 to 72%) experiencing prosthesis issues - 24.5% (range 0 to 86%) experiencing urethral issues	NA	3177 (42 studies)[6]	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the percentage of people who experience major surgical complications
Surgical complications-Minor assessed with: percentage of people experiencing major complications (Oles, 2022b) follow up: not reported	NA	The percentage varied from 9.3% (range 0% to 45.5%) experiencing donor site issues, to 24% (range 10 to 93%) experiencing urethral issues[7]	NA	4466 (52 studies)[8]	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the percentage of people who experience minor surgical complications

Table 4: All surgeries compared to no surgeries in assigned females at birth (<21 years old) with gender dysphoria

Patient or population: assigned females at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Reoperation assessed with: rate of reoperation across patients (Oles, 2022b) follow up: not reported	NA	27.6% Range (2.5% to 40%)	NA	1624 (15 studies)	⨁◯◯◯ VERY LOW[4]	We are very uncertain about the percentage of people who undergo reoperations

CI: Confidence interval
NA: Not applicable

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

1. Studies used different scales to assess satisfaction
2. The percentage was similar when the analysis was done by type of surgery and by follow up time (< 1 year vs 1 year or more). Another systematic review (Oles, 2022) also investigated this outcome, and reported a very similar percentage of satisfaction (91.8%, range 73% to 100%)
3. Percentage of patients satisfied rated down due to risk of bias and indirectness. According to the systematic review authors, several studies were judged at moderate and high risk of bias. In addition, the median of the mean age of patients included in the studies was 28 years
4. Rated down due to risk of bias, inconsistency/ imprecision, and indirectness. Even though the review authors did not assess risk of bias, these studies were included in other systematic reviews in which the authors judged several of them at high risk of bias. The studies report inconsistent results (some high and other low rates). The patients are older than the target population.
5. Results for phalloplasty. Similar results for metoidioplasty (91.3%).

6. People and studies for urethral complications. 2671 people (37 studies) for prosthesis issues, and 1548 people (22 studies) for total flap loss.
7. Percentage of wound dehiscence 9.8% (range, 2.9% to 75%), percentage of infection/ partial necrosis 10.3% (range, 0 to 45.8%), percentage of prosthesis issues 14.2% (range, 1.6 to 41.9%), percentage of incontinence 15.3% (range, 5.4% to 59.1%)
8. People and studies for infection/ partial necrosis. 2389 people (31 studies) for urethral issues, 1736 people (17 studies) for wound dehiscence, 1080 (10 studies) for prosthesis issues, 1053 people (8 studies) for donor site issues, 131 people (3 studies) for incontinence

3.3 **Effects of surgeries on assigned males at birth**: Table 5 summarizes the evidence about the effects of all surgeries on surgical outcomes among assigned at birth males. We used information from 3 systematic reviews.[16 18 19] Due to the nature of the outcomes (i.e. they can only be experienced by people who undergo surgeries), there cannot be studies comparing the outcomes between a group of people with gender dysphoria who undergo surgeries and another who does not.

Studies, therefore, assessed the outcomes among a group of people with gender dysphoria after surgery. The findings are:

- There is low certainty evidence suggesting that the percentage of people who are satisfied after vaginoplasty is high (91%)
- There is very low certainty evidence about the percentage of people who are satisfied, the rate of complications, and the rate of reoperations after chest surgery (see Table 5 for details)
- There is low certainty evidence suggesting that the percentage of people who have regret after vaginoplasty is low (2%)
- There is very low certainty evidence about the rate of complications and the rate of reoperations after vaginoplasty (see Table 5 for details)

Table 5: All surgeries compared to no surgeries in assigned males at birth (<21 years old) with gender dysphoria

Patient or population: assigned males at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	**Risk/ mean with surgery**				
Chest surgery						
Satisfaction assessed with: percentage of people who reported being satisfied (Oles 2022) Range of follow up: 12 months to 17 years	NA	Range 75% (80/107) to 95% (33/35)[1]	NA	142 (2 studies)	⨁◯◯◯ VERY LOW[2]	We are very uncertain about the percentage of people who report being satisfied
Surgical complications assessed with: rate of complications across patients (Oles 2022) Range of follow up: 2 weeks to 16 years	NA	The complication rates were: - 3.8% (range 0% to 5.5%) of capsular contracture - 2.2% of major hematoma - 2.2% of implant extrusion[3]	NA	432 (5 studies)	⨁◯◯◯ VERY LOW[2]	We are very uncertain about the rate of surgical complications
Reoperation assessed with: rate of reoperation across patients (Oles 2022) Range of follow up: Not reported	NA	8.6% Range (4.4% to 10.4%)	NA	291 (2 studies)	⨁◯◯◯ VERY LOW[2]	We are very uncertain about the rate of reoperation
Bottom surgery						

Table 5: All surgeries compared to no surgeries in assigned males at birth (<21 years old) with gender dysphoria

Patient or population: assigned males at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Satisfaction assessed with: percentage of people who reported being satisfied for overall outcomes (Bustos SS, 2021) Range of follow up: 1 week to 11.3 years	NA	91% (81% to 98%)[4]	NA	1230 (12 studies)	⨁⨁◯◯ LOW[5]	The percentage of people who report being satisfied with overall outcomes may be 91%
Regret assessed with: percentage of people who reported regret (Bustos SS, 2021) Range of follow up: 2 months to 24.1 years	NA	2% (1% to 3%)	NA	1137 (15 studies)	⨁⨁◯◯ LOW[6]	The percentage of people who report regret may be 2%
Surgical complications assessed with: rate of complications across patients (Bustos SS, 2021) Range of follow up: 3 weeks to 24.1 years	NA	The complication rates were: - 1% (95% CI, <0.1% to 2%) of fistula - 11% (95% CI, 8% to 14%) of stenosis and/or strictures - 4% (95% CI, 1% to 9%) of tissue necrosis - 3% (95% CI, 1% to 4%) of prolapse[7]	NA	4196 (42 studies)[3]	⨁◯◯◯ VERY LOW[8]	We are very uncertain about the rate of surgical complications

Table 5: All surgeries compared to no surgeries in assigned males at birth (<21 years old) with gender dysphoria

Patient or population: assigned males at birth (<21 years old) with gender dysphoria
Intervention: surgeries
Comparison: no surgeries

Outcomes	Anticipated absolute effects* (95% CI)		Relative effect (95% CI)	№ of participants (studies)	Certainty of the evidence (GRADE)	What happens
	Risk / mean with no surgery	Risk/ mean with surgery				
Reoperation assessed with: rate of reoperation across patients (Tay, 2021) Range of follow up: 6 weeks to 14.8 months	NA	One study reported a surgical revision rate of 9% (1/11 patients), and a second study reported that 13% (19/145) patients required repeat surgery due to complications.	NA	156 (2 studies)	⨁◯◯◯ VERY LOW[9]	We are very uncertain about the percentage of people who undergo reoperations

CI: Confidence interval
NA: Not applicable

GRADE Working Group grades of evidence
High certainty: We are very confident that the true effect lies close to that of the estimate of the effect
Moderate certainty: We are moderately confident in the effect estimate: The true effect is likely to be close to the estimate of the effect, but there is a possibility that it is substantially different
Low certainty: Our confidence in the effect estimate is limited: The true effect may be substantially different from the estimate of the effect
Very low certainty: We have very little confidence in the effect estimate: The true effect is likely to be substantially different from the estimate of effect

Explanations

1. Another systematic review, Sijben 2021, reported satisfaction from 3 additional studies: 82% (113/138) were satisfied or very satisfied, 93% (32/34) were happier and more satisfied with their chest, and 79% (28/36) were very satisfied with the overall cosmetic result (very low certainty of evidence due to risk of bias, imprecision, and indirectness).
2. Rated down due to risk of bias, indirectness (the included studies were not restricted to youth or young adults), and imprecision (too few participants included, not meeting optimal information size).

3. Another systematic review, Sijben 2021, reported similar ranges for rates of complication requiring reoperation from 7 studies (835 patients): capsular contraction (range 0.0-5.6%), asymmetry (3.6%), hematoma (range 0.0-2.9%), infection (range 0.0-0.9%), striae distensae (0.7%), implant rupture (0.7%), abscess (0.4%), scarring (0.0%), hypersensitivity (0.0%), and numbness (0.0%) (very low certainty of evidence due to risk of bias, imprecision, and indirectness)
4. Bustos SS *et al.* 2021 additionally reported on satisfaction for functional (87%, 95% CI 77% to 94%) and aesthetic (90%, 95% CI 84% to 94%) outcomes. Another systematic review and meta-analysis, Oles 2022b, similarly reported that 92.3% (range 23.1% to 100%) of patients (2410/2601) were satisfied after vaginoplasty (very low certainty of evidence due to risk of bias, imprecision, and indirectness).
5. Rated down due to risk of bias (the systematic review authors reported the quality of the included studies to be low to moderate using the New Castle Ottawa scale), and indirectness as the included studies were not restricted to youth or young adults. We did not rate down for imprecision or inconsistency despite high I^2 values as a satisfaction rate of 80% or above was deemed as a minimum threshold for clinical importance.
6. Rated down due to risk of bias (the systematic review authors reported the quality of the included studies to be low to moderate using the New Castle Ottawa scale), and indirectness as the included studies were not restricted to youth or young adults.
7. Another systematic review, Oles 2022b, similarly reported the percentage of patients experiencing complications from 51 studies, ranging from 2.4% to 12.0% (range 0% to 88%) for minor complications (intraoperative injury, wound dehiscence, superficial necrosis, infection, urinary issues, vaginal prolapse, stenosis, and bleeding) and 1.6% to 2.1% (range 0% to 31%) for major complications (flap/graft necrosis and infection) after genitoplasty (very low certainty of evidence due to risk of bias, imprecision, and indirectness).
8. Rated down due to risk of bias (the systematic review authors reported the quality of the included studies to be low to moderate using the New Castle Ottawa scale), imprecision and inconsistency, with wide confidence intervals and I^2 values ranging from 65.8% to 94.3%, and indirectness as the included studies were not restricted to youth or young adults.
9. Rated down due to risk of bias, indirectness (the age range of patients in the included studies was 24 to 39 years; the studies included were restricted to those that investigated the use of peritoneum in neovagina construction), and imprecision (too few participants included, not meeting optimal information size).

Results from search for studies not included in the systematic reviews: After screening 1854 records found through our searches, we found 10 eligible studies (figure 5). From these, 8 were comparative observational studies[20-27] and 2 were non- comparative[28 29]. We provide reasons for excluding studies in appendix 2.

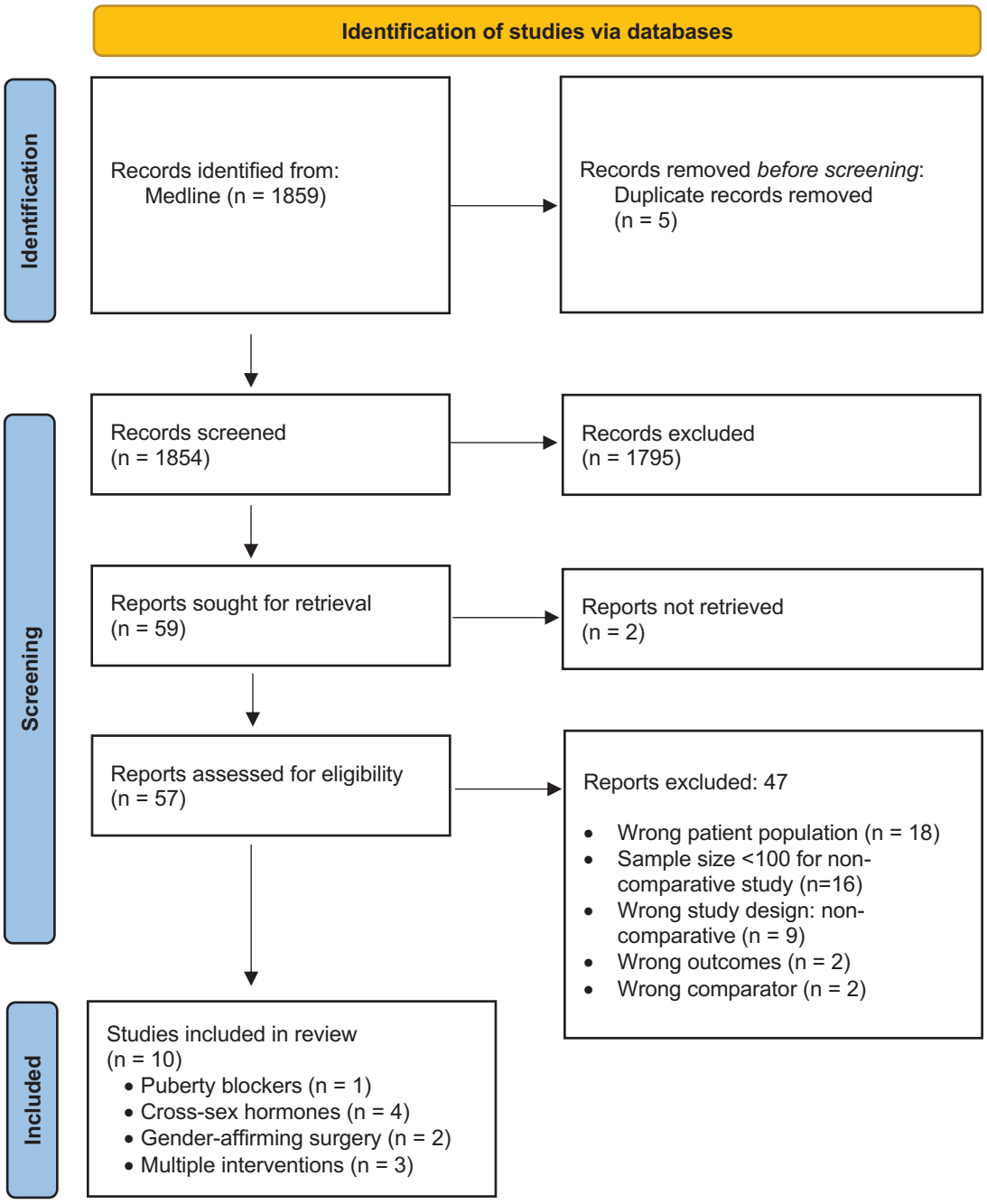


Figure 5: PRISMA flow diagram for the selection of primary studies. *From:* Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 2021;372:n71. doi: 10.1136/bmj.n71. For more information, visit: http://www.prisma-statement.org/

None of the studies were judged as likely to importantly change the conclusions obtained from the systematic reviews (Tables 6 and 7). The main limitations of the comparative studies were risk of bias concerns (Figures 6 and 7) due to confounding, classification of intervention, and missing data; as well as small sample sizes. Although non-comparative studies were at lower risk of bias, because their results were consistent with those of the included evidence, they were also judged as unlikely to change the conclusions importantly.

Table 6: Characteristics of eligible comparative observational studies

Study ID	Sample size*	Study design	Intervention	Comparator	Outcomes measured	Likely to change conclusions	Reasons
VanDerMiesen, 2020	450	Retrospective cohort study	Puberty blockers	Waiting for puberty blockers	Self-harm/ suicidality, internalizing behaviors	No	Reports a small benefit on suicidality and moderate on internalizing behaviours, but high risk of bias
Becker-Hebly, 2021	75	Prospective cohort study	1. Puberty blockers 2. Cross-sex hormones 3. Surgery	No medical intervention yet; psychosocial intervention only	Health-related quality of life	No	Critical risk of bias (missing data due to low response rate, and confounding). Reports small benefit in mean change score for mental and physical dimension QoL as compared to no medical treatment. Imprecision; the 95% CIs for mean change scores are wide.
Green, 2021	3235	Cross-sectional study	Cross-sex hormones	Would like to take cross-sex hormones	Depression, suicidality	No	Critical risk of bias, no follow up of patients (measurement of current outcomes and not adjusting for baseline)
Tordoff, 2022	84	Prospective cohort study	1. Puberty blockers 2. Cross-sex hormones	No intervention	Depression, anxiety, suicidal thoughts	No	Moderate risk of bias, small sample size
Turban, 2022	9341	Cross-sectional study	Cross-sex hormones	Desired but never accessed gender affirming hormones	Suicidal ideation, suicidal attempt	No	Critical risk of bias, no follow up of patients (measurement of current outcomes and not adjusting for baseline)
Grannis, 2021	47	Cross-sectional study	Cross-sex hormones	No intervention yet	Anxiety, depression	No	Critical risk of bias, no follow up of patients, small sample size
Fontanari, 2020	350	Cross-sectional study	1. Cross-sex hormones 2. Cross-sex hormones or surgery	1. Waiting for cross-sex hormones 2. No intervention	Anxiety, depression, gender distress	No	Critical risk of bias (confounding, self-reported classification of interventions). Online cross-sectional survery reported small benefit in anxiety and depression mean scores, and little to no effect on gender distress with cross-sex hormones and/or surgery. Non-randomized comparative study provides very low certainty evidence due to

							very serious risk of bias and serious imprecision (95% CIs include little to no effect)
Castelo-Branco, 2021	205	Cross-sectional study	Cross-sex hormones	No intervention	Anxiety, depression	No	Critical risk of bias due to confounding (non-adjusted analysis). Reported no difference observed in anxiety and depression mean scores (Symptom Checklist-90-Revised scale) between groups. Non-randomized comparative study provides low certainty evidence.

*Considered the number of participants relevant to the questions of this report, not all people included in the studies

Table 7: Characteristics of eligible non- comparative observational studies

Study ID	Sample size	Intervention	Outcomes measured	Likely to change conclusions	Reasons
Bordas, 2021	813	FtM bottom surgery	Surgical complications, satisfaction	No	Reports rate of complications (10.5%) and satisfaction (79% totally satisfied, 20% mainly satisfied) within range of effects reported by studies already included in systematic reviews. Unlikely to reduce imprecision and inconsistency within body of evidence (3177 and 1458 people, respectively) of non-comparative studies (42 and 27, respectively) to increase certainty of evidence
Elias, 2022	110	FtM top surgery	Complications	No	Reports rate of complications (16%) and revision surgery (5%), which is consistent with the rates reported in the studies included. Unlikely to increase the certainty of evidence

Figure 6: Risk of bias judgements for comparative studies

Study ID	Intervention	Confounding	Classification of the intervention	Deviations from intended interventions	Missing data	Measurement of outcome	Overall
Becker-Hebly, 2021	Puberty blockers, cross-sex hormones, or surgery						CRITICAL
Castelo-Branco, 2021	Cross-sex hormones						CRITICAL
Fontanari, 2020	Cross-sex hormones, cross-sex hormones or surgery						CRITICAL
Grannis, 2021	Cross-sex hormones						CRITICAL
Green, 2021	Cross-sex hormones						CRITICAL
Tordoff, 2022	Puberty blockers, cross-sex hormones						MODERATE
Turban, 2022	Cross-sex hormones						CRITICAL
Van Der Miesen, 2020	Puberty blockers						SERIOUS

Figure legend:

Low

Moderate

Serious

Critical

Unclear

Figure 7: Risk of bias judgements for non-comparative studies

Study ID	Intervention	Representativeness of sample	Classification of intervention	Deviation from intended interventions	Missing data	Measurement of outcome	Overall
Bordas, 2021	FtM bottom surgery						LOW
Elias, 2022	FtM top surgery						MODERATE

Figure legend:

Low

Moderate

High

Effects of gender affirming therapies in people with gender dysphoria: evaluation of the best available evidence. Dr. Romina Brignardello-Petersen and Dr. Wojtek Wiercioch; Results; May 16, 2022

References

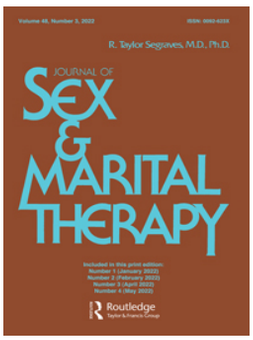
1. Ramos GGF, Mengai ACS, Daltro CAT, et al. Systematic Review: Puberty suppression with GnRH analogues in adolescents with gender incongruity. *Journal of endocrinological investigation* 2021;44(6):1151-58. doi: https://dx.doi.org/10.1007/s40618-020-01449-5
2. Rew L, Young CC, Monge M, et al. Review: Puberty blockers for transgender and gender diverse youth-a critical review of the literature. *Child and adolescent mental health* 2021;26(1):3-14. doi: https://dx.doi.org/10.1111/camh.12437
3. Excellence NIfHaC. Evidence review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria, 2020.
4. Quality AfHRa. Topic Brief: Treatments for Gender Dysphoria in Transgender Youth, 2021.
5. Baker KE, Wilson LM, Sharma R, et al. Hormone Therapy, Mental Health, and Quality of Life Among Transgender People: A Systematic Review. *Journal of the Endocrine Society* 2021;5(4):bvab011. doi: 10.1210/jendso/bvab011
6. Fledderus AC, Gout HA, Ogilvie AC, et al. Breast malignancy in female-to-male transsexuals: systematic review, case report, and recommendations for screening. *Breast (Edinburgh, Scotland)* 2020;53(9213011):92-100. doi: https://dx.doi.org/10.1016/j.breast.2020.06.008
7. Haupt C, Henke M, Kutschmar A, et al. Antiandrogen or estradiol treatment or both during hormone therapy in transitioning transgender women. *The Cochrane database of systematic reviews* 2020;11:CD013138. doi: 10.1002/14651858.CD013138.pub2
8. Karalexi MA, Georgakis MK, Dimitriou NG, et al. Gender-affirming hormone treatment and cognitive function in transgender young adults: a systematic review and meta-analysis. *Psychoneuroendocrinology* 2020;119:104721. doi: 10.1016/j.psyneuen.2020.104721
9. Kotamarti VS, Greige N, Heiman AJ, et al. Risk for Venous Thromboembolism in Transgender Patients Undergoing Cross-Sex Hormone Treatment: A Systematic Review. *The journal of sexual medicine* 2021 doi: 10.1016/j.jsxm.2021.04.006
10. Mattawanon N, Charoenkwan K, Tangpricha V. Sexual Dysfunction in Transgender People: A Systematic Review. *The Urologic clinics of North America* 2021;48(4):437-60. doi: 10.1016/j.ucl.2021.06.004
11. Totaro M, Palazzi S, Castellini C, et al. Risk of Venous Thromboembolism in Transgender People Undergoing Hormone Feminizing Therapy: A Prevalence Meta-Analysis and Meta-Regression Study. *Frontiers in endocrinology* 2021;12:741866. doi: 10.3389/fendo.2021.741866
12. Excellence NIfHaC. Evidence review: Gender-affirming hormones for children and adolescents with gender dysphoria., 2020.
13. Eftekhar Ardebili M, Janani L, Khazaei Z, et al. Quality of life in people with transsexuality after surgery: a systematic review and meta-analysis. *Health and quality of life outcomes* 2020;18(1):264. doi: 10.1186/s12955-020-01510-0
14. Bustos VP, Bustos SS, Mascaro A, et al. Regret after Gender-affirmation Surgery: A Systematic Review and Meta-analysis of Prevalence. *Plastic and reconstructive surgery Global open* 2021;9(3):e3477. doi: 10.1097/GOX.0000000000003477
15. Oles N, Darrach H, Landford W, et al. Gender Affirming Surgery: A Comprehensive, Systematic Review of All Peer-reviewed Literature and Methods of Assessing Patient-centered Outcomes (Part 2: Genital Reconstruction). *Annals of surgery* 2022;275(1):e67-e74. doi: 10.1097/SLA.0000000000004717
16. Oles N, Darrach H, Landford W, et al. Gender Affirming Surgery: A Comprehensive, Systematic Review of All Peer-Reviewed Literature and Methods of Assessing Patient-Centered Outcomes (Part 1: Breast/Chest, Face, and Voice). *Annals of surgery* 2022 doi: 10.1097/SLA.0000000000004728
17. Bustos VP, Bustos SS, Mascaro A, et al. Transgender and Gender-nonbinary Patient Satisfaction after Transmasculine Chest Surgery. *Plastic and reconstructive surgery Global open* 2021;9(3):e3479. doi: 10.1097/GOX.0000000000003479
18. Bustos SS, Bustos VP, Mascaro A, et al. Complications and Patient-reported Outcomes in Transfemale Vaginoplasty: An Updated Systematic Review and Meta-analysis. *Plastic and reconstructive surgery Global open* 2021;9(3):e3510. doi: 10.1097/GOX.0000000000003510
19. Tay YT, Lo CH. Use of peritoneum in neovagina construction in gender-affirming surgery: A systematic review. *ANZ journal of surgery* 2021 doi: 10.1111/ans.17147

20. Becker-Hebly I, Fahrenkrug S, Campion F, et al. Psychosocial health in adolescents and young adults with gender dysphoria before and after gender-affirming medical interventions: a descriptive study from the Hamburg Gender Identity Service. *European child & adolescent psychiatry* 2021;30(11):1755-67. doi: https://dx.doi.org/10.1007/s00787-020-01640-2
21. Castelo-Branco C, RiberaTorres L, Gomez-Gil E, et al. Psychopathological symptoms in Spanish subjects with gender dysphoria. A cross-sectional study. *Gynecological endocrinology : the official journal of the International Society of Gynecological Endocrinology* 2021;37(6):534-40. doi: https://dx.doi.org/10.1080/09513590.2021.1913113
22. Fontanari AMV, Vilanova F, Schneider MA, et al. Gender Affirmation Is Associated with Transgender and Gender Nonbinary Youth Mental Health Improvement. *LGBT health* 2020;7(5):237-47. doi: https://dx.doi.org/10.1089/lgbt.2019.0046
23. Grannis C, Leibowitz SF, Gahn S, et al. Testosterone treatment, internalizing symptoms, and body image dissatisfaction in transgender boys. *Psychoneuroendocrinology* 2021;132(7612148, qgc):105358. doi: https://dx.doi.org/10.1016/j.psyneuen.2021.105358
24. Green AE, DeChants JP, Price MN, et al. Association of Gender-Affirming Hormone Therapy With Depression, Thoughts of Suicide, and Attempted Suicide Among Transgender and Nonbinary Youth. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine* 2022;70(4):643-49. doi: https://dx.doi.org/10.1016/j.jadohealth.2021.10.036
25. Tordoff DM, Wanta JW, Collin A, et al. Mental Health Outcomes in Transgender and Nonbinary Youths Receiving Gender-Affirming Care. *JAMA network open* 2022;5(2):e220978. doi: https://dx.doi.org/10.1001/jamanetworkopen.2022.0978
26. Turban JL, King D, Kobe J, et al. Access to gender-affirming hormones during adolescence and mental health outcomes among transgender adults. *PloS one* 2022;17(1):e0261039. doi: https://dx.doi.org/10.1371/journal.pone.0261039
27. van der Miesen AIR, Steensma TD, de Vries ALC, et al. Psychological Functioning in Transgender Adolescents Before and After Gender-Affirmative Care Compared With Cisgender General Population Peers. *The Journal of adolescent health : official publication of the Society for Adolescent Medicine* 2020;66(6):699-704. doi: https://dx.doi.org/10.1016/j.jadohealth.2019.12.018
28. Bordas N, Stojanovic B, Bizic M, et al. Metoidioplasty: Surgical Options and Outcomes in 813 Cases. *Frontiers in endocrinology* 2021;12(101555782):760284. doi: https://dx.doi.org/10.3389/fendo.2021.760284
29. Elias N, Rysin R, Kwartin S, et al. Breaking the Binary: The Approach to Chest Masculinizing Gender-Affirming Surgery in Trangender Men. *The Israel Medical Association journal : IMAJ* 2022;24(1):20-24.

ID	Study	Reason
#534	Abu-Ghname 2020	Wrong population: non transgender men
#434	Aires 2022	Wrong interventions: Other type of surgery: glottoplasty
#514	Angus 2021	Wrong outcomes: It does not include any outcome of interest. Includes: serum total testosterone concentration, body fat redistribution, breast development, and facial/body hair reduction
#318	Baddredine 2022	Wrong intervention. Continuing vs stopping estrogen during perioperative period of vaginoplasty
#40	Baram 2019	Wrong outcomes: only clinical outcomes are sperm count, testicular histology, hormone levels, etc.
#145	Barcelos 2022	Wrong outcomes: sexual satisfaction, desire, and function outcomes only
#60	Boczar 2021	No outcome data
#386	Bouman 2014	Wrong population: unclear that more than 80% are transgender
#208	Bustos 2021	Wrong intervention: niple areola reconstruction
#54	Connelly 2021	Wrong outcomes: Blood pressure
#43	Coon 2022	Wrong intervention: facial gender surgery
#34	D'Angelo 2018	Wrong design: narrative review
#165	Delgado-Ruiz 2019	Wrong outcomes: bone density
#355	Escandon 2022	Other type of surgery: facial surgery
#129	Fighera 2019	Wrong outcomes: bone mass
#597	Hembree 2017	Practice guideline, does not report the methods/ results of the systematic review in details
#120	Kakadekar 2021	Wrong outcomes: histological findings
#451	Kennedy 2021	Wrong intervention: self administered hormones
#375	Kloer 2021	Wrong outcomes: sexual health and satisfaction outcomes only
#439	Kovar 2019	More than 20% participants did not have gender dysphoria
#297	Kristensen 2021	Wrong outcomes: agression and hostility
#637	Leclere 2015	Wrong design: commentary of a systematic review
#293	Miranda 2021	Published in abstract format only
#624	Morrison 2016	Wrong intervention: facial feminization surgery
#270	Narayan 2021	Wrong design: narrative review
#119	Nolan 2019	Wrong intervention: phonosurgery
#167	Patel 2021	Wrong intervention: facial hair transplantation
#287	Ray 2020	Wrong population: cisgender is the population of interest, transgender included as indirect evidence and not in a systematic manner
#518	Rozga 2020	Published in abstract format only
#265	Sariyaka 2017	Wrong population: More than 20% participants did not have gender dysphoria
#35	Sayegh 2019	Wrong intervention: facial masculinization surgery
#124	Schwarz 2017	Wrong intervention: laryngeal surgery

#97	Siringo 2021	Wrong intervention: facial feminization surgery
#253	Song 2016	Wrong intervention: phonosurgery
#250	Song 2017	Wrong intervention: phonosurgery
#104	Spanos 2020	Wrong outcomes: lean mass, fat mass or insulin resistance
#257	Therattil 2017	Wrong intervention: thyroid chartilage reduction surgery
#328	Tirrell 2022	Wrong intervention: facial feminization surgery
#676	Traish 2010	Wrong design: narrative review
#279	VanDamme 2017	Wrong intervention: voice pitch raising surgery
#171	Vellho 2017	Wrong outcomes: BMI, blood pressure, hematocrit, hemoglobin, lipid profile, and liver enzymes
#112	Wilson 2020	Wrong outcomes: prolactine related outcomes (levels, hyperprolactinemia, prolactinoma)
#245	Worth 2018	Unable to access full text
#122	Ziegler 2018	Wrong outcomes: voice parameters and satisfaction with voice
#499	Zucker 2021	Unable to access full text

ID	Study	Reason
#1458	Al-Tamimi 2019	Wrong patient population
#287	Al-Tamimi 2020	Wrong study design: non comparative
#403	Alcon 2021	Wrong study design: non comparative
#214	Aldridge 2021	Wrong study design: non comparative
#54	Almazan 2021	Wrong patient population
#1387	Boas 2019	Wrong patient population
#1323	Branstrom 2020	Wrong patient population
#1447	Breidenstein 2019	Wrong study design: non comparative
#114	Briles 2022	Insufficient Sample Size <100
#1804	Butler 2019	Wrong patient population
#716	Carmichael 2021	Wrong study design: non comparative
#622	Cocchetti 2021	Wrong outcomes
#1067	Coon 2020	Wrong patient population
#1835	Cristofari 2019	Wrong patient population
#1486	Cuccolo 2019	Wrong patient population
#1276	deBlok 2020	Wrong patient population
#577	deRooij 2021	Wrong patient population
#1625	DeWolf 2019	Wrong patient population
#1759	Djordjevic 2019	Wrong patient population
#244	Falcone 2020	Insufficient Sample Size <100
#258	FosterSkewis 2021	Wrong comparator
#1583	Gallagher 2019	Wrong patient population
#139	Gumussoy 2022	Wrong study design: non comparative
#515	Hisle-Gorman 2021	Wrong study design: non comparative
#350	Hougen 2021	Insufficient Sample Size <100
#1007	Meyer 2020	Wrong study design: non comparative
#499	Miller 2021	Wrong patient population
#621	Mullins 2021	Wrong study design: non comparative
#1653	Naeimi 2019	Insufficient Sample Size <100
#1691	Namba 2019	Insufficient Sample Size <100
#1770	Neuville 2019	Insufficient Sample Size <100
#623	Neuville 2021	Insufficient Sample Size <100
#644	Nieder 2021	Insufficient Sample Size <100
#1624	Nikkels 2019	Wrong patient population
#353	Opsomer 2021	Wrong patient population
#1306	Papadopulos 2020	Wrong comparator
#640	Papadopulos 2021	Insufficient Sample Size <100
#1472	Pigot 2019	Wrong patient population
#899	Pigot 2020	Insufficient Sample Size <100
#1212	Segev-Becker 2020	Insufficient Sample Size <100
#1351	Staples 2020	Wrong outcomes
#645	Staud 2021	Insufficient Sample Size <100
#864	Terrier 2020	Insufficient Sample Size <100
#1083	vanderSluis 2020	Insufficient Sample Size <100



Routledge
Taylor & Francis Group

Journal of Sex & Marital Therapy

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/usmt20

Reconsidering Informed Consent for Trans-Identified Children, Adolescents, and Young Adults

Stephen B. Levine, E. Abbruzzese & Julia M. Mason

To cite this article: Stephen B. Levine, E. Abbruzzese & Julia M. Mason (2022): Reconsidering Informed Consent for Trans-Identified Children, Adolescents, and Young Adults, Journal of Sex & Marital Therapy, DOI: 10.1080/0092623X.2022.2046221

To link to this article: https://doi.org/10.1080/0092623X.2022.2046221

Published online: 17 Mar 2022.

JOURNAL OF SEX & MARITAL THERAPY
https://doi.org/10.1080/0092623X.2022.2046221

Routledge
Taylor & Francis Group

REVIEW

OPEN ACCESS Check for updates

Reconsidering Informed Consent for Trans-Identified Children, Adolescents, and Young Adults

Stephen B. Levine[a], E. Abbruzzese[b] and Julia M. Mason[c]

[a]Department of Psychiatry, Case Western Reserve University, Cleveland, OH, USA; [b]Society for Evidence Based Gender Medicine (SEGM), Twin Falls, ID, USA; [c]Department of Pediatrics, Gresham, OR, USA

ABSTRACT

In less than a decade, the western world has witnessed an unprecedented rise in the numbers of children and adolescents seeking gender transition. Despite the precedent of years of gender-affirmative care, the social, medical and surgical interventions are still based on very low-quality evidence. The many risks of these interventions, including medicalizing a temporary adolescent identity, have come into a clearer focus through an awareness of detransitioners. The risks of gender-affirmative care are ethically managed through a properly conducted informed consent process. Its elements—deliberate sharing of the hoped-for benefits, known risks and long-term outcomes, and alternative treatments—must be delivered in a manner that promotes comprehension. The process is limited by: erroneous professional assumptions; poor quality of the initial evaluations; and inaccurate and incomplete information shared with patients and their parents. We discuss data on suicide and present the limitations of the Dutch studies that have been the basis for interventions. Beliefs about gender-affirmative care need to be separated from the established facts. A proper informed consent processes can both prepare parents and patients for the difficult choices that they must make and can ease professionals' ethical tensions. Even when properly accomplished, however, some clinical circumstances exist that remain quite uncertain.

KEYWORDS

Informed consent; ethics; gender dysphoria; gender identity; detransition

Introduction

Reconsideration of the meanings, purposes, indications, and processes of informed consent for transgender-identified youth is urgently needed. Parents of gender atypical children are considering social transition as early as preschool or grade school. Parents of preteens and teens are considering supporting their children's wishes to present in a new gender, take puberty blockers, cross-sex hormones, and plan for surgical alterations. College-aged youth are declaring new identities for the first time and obtaining hormones and surgery without their parents' knowledge.

When uncertain parents of children and teens consult their primary care providers, they are usually referred to specialty gender services. Parents and referring clinicians assume that specialists with "gender expertise" will undertake a thorough evaluation. However, the evaluations preceding the recommendation for gender transition are often surprisingly brief (Anderson & Edwards-Leeper, 2021) and typically lead to a recommendation for hormones and surgery, known as *gender-affirmative* treatment.

CONTACT Stephen B. Levine sbl2@case.edu Department of Psychiatry, Case Western Reserve University, 23425 Commerce Park #104, Cleveland, OH44106-7078, USA.

Despite the widely recognized deficiencies in the evidence supporting gender-affirmative interventions (National Institute for Health & Care Excellence, 2020a; 2020b), the process of obtaining informed consent from patients and their families has no established standard. There is no consensus about the requisite elements of evaluations, nor is there unanimity about how informed consent processes should be conducted (Byne et al., 2012). These two matters are inconsistent from practitioner to practitioner, clinic to clinic, and country to country.

Social transition, hormonal interventions, and surgery have profound implications for the course of the lives of young patients and their families. It is incumbent upon professionals that these consequences be thoroughly, patiently clarified over time prior to undertaking any element of transition. The informed consent process does not preclude transition; it merely educates the family about the state of the science underpinning the decision to transition. Social transition, hormones, and surgeries are unproven in a strict scientific sense, and as such, to be ethical, require a thorough and fully informed consent process.

Ethical Concerns About Inadequate Informed Consent

The concept of informed consent in medicine has roots in both ethical theory and law. The ethical foundation is centered in the principles of beneficence, justice, and respect for autonomy, while the legal issues have to do with questions of malpractice (Katz et al., 2016).

Patients consenting to treatment must meet age-based and decisional capacity requirements (Katz et al., 2016). Minors less than the age of consent participate in decision-making by providing *assent*—an agreement with the intervention. The limited maturational cognitive capacities of minors are the key reason why parents serve as the ethical and legal surrogates for medical decision-making, tasked with signing an informed consent document (Grootens-Wiegers, Hein, van den Broek, & de Vries, 2017).

The informed consent process consists of three main elements: a disclosure of information about the nature of the condition and the proposed treatment and its alternatives; an assessment of patient and caregiver understanding of the information and capacity for medical decision-making; and obtaining the signatures that signify informed consent has been obtained (Katz et al., 2016). The current expectation that clinicians and institutions are required to thoroughly inform their patients about the benefits, risks, and uncertainties of a particular treatment, as well as about alternatives, has a long legal history in the United States (Lynch, Joffe, & Feldman, 2018).

Ethical concerns about inadequate informed consent for trans-identified youth have several potentially problematic sources, including *erroneous assumptions* held by professionals; *poor quality of the evaluation process*; and *incomplete and inaccurate information* that the patients and family members are given.

These concerns are amplified by the *dramatic growth* in demand for youth gender transition witnessed in the last several years that has led to a perfunctory informed consent process. A rushed process does not allow for a proper discussion of not only the benefits, but the profound risks and uncertainties associated with gender transition, especially when gender transition is undertaken before mature adulthood.

a. *Dramatic growth in demand for services threatens true informed consent*

Gender identity variations were thought to be extremely rare a generation ago. While the incidence in youth had not been officially estimated, in adults it was 2-14 per 100,000 (American Psychiatric Association, 2013, p. 454). However, around 2006, the incidence among youth began to rise, with a dramatic increase observed in 2015 (Aitken et al., 2015, de Graaf, Giovanardi, Zitz, & Carmichael, 2018). Currently, 2-9% of U.S. high school students now identify as transgender, while in colleges, 3% of males and 5% of females identify as gender-diverse (American College Health Association, 2021; Johns et al., 2019; Kidd et al., 2021).

Whereas previously most of the affected individuals identified as the opposite sex, there is now a growing trend toward identifying as *nonbinary*: neither male nor female or both male and female (Chew et al., 2020). A recent study reported that the majority of transgender-identifying youth (63%) now have a non-binary identity (Green, DeChants, Price, & Davis, 2021). Although the incidence of natal males asserting a trans identity in adolescence has significantly increased, the dramatic increase is driven primarily by the increase in natal females requesting services (Zucker, 2017). Many suffer from significant comorbid mental health disorders, have neurocognitive difficulties such as ADHD or autism or have a history of trauma (Becerra-Culqui et al., 2018; Kozlowska, McClure, et al., 2021).

The increase in rates of transgender identification is reflected in the numbers of youth seeking help from medical professionals. For example, according to data reported by the Tavistock gender clinic in the UK, in 2009, there were 51 requests for services (de Graaf et al., 2018); in 2019-2020, 2728 referrals were recorded—a 53-fold increase in just over a decade (Tavistock & Portman NHS Foundation Trust, 2020). The growing number of urban transgender health centers that have arisen in recent years (HRC, n.d.) reflects the increased demand for gender-related medical care among young people in North America Australia, and Europe.

This unprecedented increase has created pressure on institutions and practitioners to rapidly evaluate these youth and make recommendations about treatment. To respond to growing demand, an innovative *informed consent model of care* has been developed. Under this model, mental health evaluations are not required, and hormones can be provided after just one visit following the collection of a patient's or guardian's consent signature (Schulz, 2018). The provision of transition services under this model of care is available not just to those over 18, but for younger patients as well (Planned Parenthood League of Massachusetts, n.d.).

Although following the informed consent model of care for hormones and surgeries for youth may diminish clinicians' ethical or moral unease (Vrouenraets et al., 2020), we believe this model is the antithesis of true informed consent, as it jeopardizes the ethical foundation of patient autonomy. Autonomy is not respected when patients consenting to the treatment do not have an accurate understanding of the risks, benefits, and alternatives.

b. *Assumptions held by professionals influence the integrity of the informed consent process*

Gender dysphoric children and teens can intensely occupy the belief that their lives will be immensely improved by transition. Clinicians who have embraced the gender-affirmative model of care operate on the assumption that children and teens know best what they need to be happy and productive (Ehrensaft, 2017). These professionals, responding to the youths' passionate pleas, see their role as validating the young person's fervent wishes for hormones and surgery and clearing the path for gender transition. In doing so, they privilege the ethical principle of respect for patient autonomy (Clark & Virani, 2021) over their obligations for beneficence and non-maleficence.

Many of the gender-affirmative clinicians subscribe to the theory of *minority stress* – the supposition that the frequently co-occurring psychiatric symptoms of gender dysphoric individuals are a result of prejudice and discrimination brought about by gender non-conformity (Rood et al., 2016; Zucker, 2019), and that gender transition will ameliorate these symptoms. Some even claim that gender-affirmative care will successfully treat not only depression and anxiety but will also resolve neurocognitive deficits frequently present in gender dysphoric individuals (Turban, 2018; Turban, King, Carswell, & Keuroghlian, 2020; Turban & van Schalkwyk, 2018). These latter assertions have proven controversial even among the proponents of gender-affirmative interventions (Strang et al., 2018; van der Miesen, Cohen-Kettenis, & de Vries, 2018). The minority stress theory as the sole explanatory mechanism for co-occurring mental health illness has also been questioned in light of the evidence that psychiatric symptoms frequently pre-date the onset of gender dysphoria (Bechard, VanderLaan, Wood, Wasserman, & Zucker, 2017; Kaltiala-Heino, Sumia, Työläjärvi, & Lindberg, 2015; Kozlowska, Chudleigh, McClure, Maguire,

& Ambler, 2021). Other clinicians recognize the limits of gender-affirmative care and are aware that youth with underlying psychiatric issues are likely to continue to struggle post-transition (Kaltiala, Heino, Työläjärvi, & Suomalainen, 2020), but, unaware of alternative approaches such as gender-exploratory psychotherapy or watchful waiting (Bonfatto & Crasnow, 2018; Churcher Clarke & Spiliadis, 2019; Spiliadis, 2019), these well-meaning professionals continue to treat youth with gender-affirmative interventions despite lingering doubts.

It is common for gender-affirmative specialists to erroneously believe that gender-affirmative interventions are a *standard of care* (Malone, D'Angelo, Beck, Mason, & Evans, 2021; Malone, Hruz, Mason, Beck, et al:, 2021). Despite the increasingly widespread professional beliefs in the safety and efficacy of pediatric gender transition, and the endorsement of this treatment pathway by a number of professional medical societies, the best available evidence suggests that the benefits of gender-affirmative interventions are of very low certainty (Clayton et al., 2021; National Institute for Health & Care Excellence, 2020a; 2020b) and must be carefully weighed against the health risks to fertility, bone, and cardiovascular health (Alzahrani et al., 2019; Biggs, 2021; Getahun et al., 2018; Hembree et al., 2017; Nota et al., 2019). Recently, emphasis has also been placed on psychosocial risks and as yet unknown medical risks (Malone, D'Angelo, et al., 2021).

Five scientific observations question and refute the assumption that an individual's experience of incongruence of sex and gender identity is best addressed by supporting the newly assumed gender identity with psychosocial and medical interventions.

1. The most foundational aspect of the diagnoses of "gender dysphoria" (DSM-5) and "gender incongruence" (ICD-11), requisite for the provision of medical treatment, is in flux, as professionals disagree on whether the presence of distress is a key diagnostic criterion, as stated in the DSM-5, or is irrelevant, as is the case according to the latest ICD-11 criteria (American Psychiatric Association, 2013; World Health Organization, 2019). Further, these diagnoses have never been properly field-tested (de Vries et al., 2021).
2. There are no randomized controlled studies demonstrating the superiority of various affirmative interventions compared to alternatives. There isn't even agreement about which outcome measures would be ideal in such studies.
3. There are few long-term follow-up studies of various interventions using predetermined outcome measures at designated intervals. Studies that have been conducted are, at best, inconsistent. Higher quality studies with longer-follow-up fail to demonstrate durable positive impacts on mental health (Bränström & Pachankis, 2020a; 2020b).
4. Rates of post-transition desistance, increased mental suffering, increased incidence of physical illness, educational failure, vocational inconstancy, and social isolation have not been established.
5. Numerous cross-sectional and prospective studies of transgender adults consistently demonstrate a high prevalence of serious mental health and social problems as well as suicide (Asscheman et al., 2011; Dhejne et al., 2011). Controversies about how to deal with trans-identified youth must consider the well described vulnerabilities of transgender adults.

It is equally important to realize that to date, research about alternative approaches, such as psychotherapy or watchful waiting, shares the scientific limitations of the research of more invasive interventions: there are no control groups, nor is there systematic follow-up at predetermined intervals with predetermined means of measurement (Bonfatto & Crasnow, 2018; Churcher Clarke & Spiliadis, 2019; Spiliadis, 2019). Parents and patients need to be informed of this as well.

Perhaps the single most problematic assumption held by some gender clinicians is that the young patients have simply been "born in the wrong body." This assumption seemingly frees clinicians from having to contend with the ethical dilemmas of recommending body-altering

interventions that are based on very low-quality evidence. Despite the principle of development that biology, psychosocial factors, and culture generate behavior, these clinicians may believe that atypical genders are created by biology. This reductionistic approach has been criticized repeatedly (Kendler, 2019).

While the origins of childhood or adolescent onset of gender incongruence have not yet been fully elucidated, brain studies of increasing technical sophistication have yet to demonstrate a distinct structure or pattern that accounts for an atypical gender identity, after statistically controlling for sexual orientation and exposure to exogenous hormones (Frigerio, Ballerini, & Valdés Hernández, 2021). Twin studies also demonstrate that while biology plays a role in one's experience of "gender incongruence," it is far from deterministic (Diamond, 2013).

A growing number of clinicians and researchers are noting that the dramatic rise of teens declaring a trans identity appears to be, at least in part, a result of peer influence (Anderson, 2022; Hutchinson, Midgen, & Spiliadis, 2020 Littman, 2018; Littman, 2020; Zucker, 2019). Some have noted yet another influx of trans-identified youth emerging during the COVID lockdowns, and have hypothesized that increased isolation coupled with heavy internet exposure may be responsible (Anderson, 2022). While the research into the phenomenon of social influence as a contributor to trans identification of youth is still in its infancy, the possibility that clinicians are providing treatments with permanent consequences to address what may be transient identities in youth poses a serious ethical dilemma.

c. *Poor evaluations*

There is a growing recognition that rapid evaluations which disregard factors contributing to the development of gender dysphoria in youth are problematic. In November 2021, two leaders of the World Professional Organization for Transgender Health (WPATH) warned the medical community that the "The mental health establishment is failing trans kids" (Anderson & Edwards-Leeper, 2021). Frequently, evaluations provided by gender clinicians may only ascertain the diagnosis of *gender dysphoria* (DSM-5) or its ICD-11 counterpart *gender incongruence*, and screen for conspicuous mental illness prior to recommending hormones and surgeries. These limited, abbreviated evaluations overlook, and as a result fail to address, the relevant issue of the forces that may have influenced the young person's current gender identity.

Confirming the young person's self-diagnosis of gender dysphoria or gender incongruence is easy. Clarifying the developmental forces that have influenced it and determining an appropriate intervention are not. Contextualizing these forces involves an understanding of child and adolescent developmental processes, childhood adversity, co-existing physical and cognitive disadvantages, unfortunate parental or family circumstances (Levine, 2021), as well as the role of social influence (Anderson, 2022; Anderson & Edwards-Leeper, 2021; Littman, 2018; 2021).

The poor quality of mental health evaluations has been a point of significant discontent for a growing number of parents of gender dysphoric youth. Increasingly, parents have formed dozens of support groups in North America, Europe, Australia and New Zealand, united in their objections to the idea that the best or the only treatment for their gender dysphoric children is affirmation (Genspect, 2021). These distressed parents, recognizing that their son or daughter may eventually decide to present to others as a trans person, want a psychotherapeutic investigation to understand what contributed to the development of this identity and an exploration of noninvasive treatment options. Frequently, they cannot find anyone in their community who does not recommend immediate affirmation.

The American Academy of Pediatrics' Committee of Bioethics recognizes that "parents…are better situated than others to understand the unique needs of their children and to make appropriate, caring decisions regarding their children's health care" (Katz et al., 2016). The plight of the families unable to find specialists capable of conducting thorough evaluations draws attention to the widespread acceptance of medical interventions for gender-dysphoric youth as the first line of treatment. The problem is that such care has been established through precedent rather

than through scientific demonstrations of its efficacy. We contend that parents and patients have a right to know this, and that it is the professionals' responsibility and obligation to inform them of the state of knowledge in this arena of care.

d. *Incorrect information shared*

In sharing the information with patients and families, two key areas of uncertainty must be emphasized. The first one is the uncertain permanence of a child's or an adolescent's gender identity (Littman, 2021; Ristori & Steensma, 2016; Singh, Bradley, & Zucker, 2021; Vandenbussche, 2021; Zucker, 2017). The second is the uncertain long-term physical and psychological health outcomes of gender transition (National Institute for Health & Care Excellence, 2020a; 2020b). Unfortunately, gender specialists are frequently unfamiliar with, or discount the significance of, the research in support of these two concepts. As a result, the informed consent process rarely adequately discloses this information to patients and their families.

Problematically, it is common for gender clinicians to emphasize the risk of suicide if a young person's wish to transition gender is not immediately fulfilled. There is a significant amount of misinformation surrounding the question of suicidality of trans-identified youth (Biggs, 2022). Providers of gender-affirmative care should be careful not to unwittingly propagate misinformation regarding suicide to parents and youths. They should also be reminded that any conversations about suicide should be handled with great care, due to its socially contagious nature (Bridge et al., 2020; HHS, 2021).

i. High Rate of desistance/natural resolution of gender dysphoria in children is not disclosed

There have been eleven research studies to date indicating a high rate of resolution of gender incongruence in children by late adolescence or young adulthood without medical interventions (Cantor, 2020; Ristori & Steensma, 2016; Singh et al., 2021). An attempt has been made to discount the applicability of this research, suggesting that the studies were based on merely gender non-conforming, rather than truly gender-dysphoric, children (Temple Newhook et al., 2018). However, a reanalysis of the data prompted by this critique confirmed the initial finding: Among children meeting the diagnostic criteria for "Gender Identity Disorder" in DSM-IV (currently "Gender Dysphoria in DSM-5), 67% were no longer gender dysphoric as adults; the rate of natural resolution for gender dysphoria was 93% for children whose gender dysphoria was significant but subthreshold for the DSM diagnosis (Zucker, et al., 2018). It should be noted that high resolution of childhood-onset gender dysphoria had been recorded before the practice of social transition of young children was endorsed by the American Academy of Pediatrics (Rafferty et al., 2018). It is possible that social transition will predispose a young person to persistence of transgender identity long-term (Zucker, 2020).

The information regarding the resolution of gender dysphoria among those with adolescent-onset gender dysphoria, which is currently the predominant presentation, is less clear. A growing body of evidence suggests that for many teens and young adults, a post-pubertal onset of transgender identification can be a transient phase of identity exploration, rather than a permanent identity, as evidenced by a growing number of young detransitioners (Entwistle, 2020; Littman, 2021; Vandenbussche, 2021). Previously, the rate of detransition and regret was reported to be very low, although these estimates suffered from significant limitations and were likely undercounting true regret (D'Angelo, 2018). However, in the last several years since gender-affirmative care has become popularized, the rate of detransition appears to be accelerating.

According to a recent study from a UK adult gender clinic, 6.9% of those treated with gender-affirmative interventions detransitioned within only 16 months of starting treatment, and another 3.4% had a pattern of care suggestive of detransition, yielding a rate of probable detransition in excess of 10%. Another 21.7% of patients disengaged from the clinic without completing

their treatment plan (Hall, Mitchell, & Sachdeva, 2021). While some of these individuals later reengaged with the gender service, the authors concluded, "detransitioning might be more frequent than previously reported." Another study from a UK primary care practice found that 12.2% of those who had started hormonal treatments either detransitioned or documented regret, while the total of 20% stopped the treatments for a wider range of reasons. The mean age of their presentation with gender dysphoria was 20, and the patients had been taking gender-affirming hormones for the average 5 years (17 months-10 years) prior to discontinuing.

Comparing these much higher rates of treatment discontinuation and detransition to the significantly lower rates reported by the older studies, the researchers noted: "Thus, the detransition rate found in this population is novel and questions may be raised about the phenomenon of overdiagnosis, overtreatment, or iatrogenic harm as found in other medical fields" (Boyd, Hackett, & Bewley, 2022 p.15). Indeed, given that regret may take up to 8-11 years to materialize (Dhejne, Öberg, Arver, & Landén, 2014; Wiepjes et al., 2018), many more detransitioners are likely to emerge in the coming years. Detransitioner research is still in its infancy, but two recently published studies examining detransitioner experiences report that detransitioners from the recently-transitioning cohorts feel they had been rushed to medical gender-affirmative interventions with irreversible effects, often without the benefit of appropriate, or in some instances any, psychologic exploration (Littman, 2021; Vandenbussche, 2021).

Clinicians should also disclose to patients and parents that there is no test which can accurately predict who will persist in their transgender identification upon reaching mature adulthood (Ristori & Steensma, 2016). Families should be made aware that a period of strong cross-sex identification in childhood is commonly associated with future homosexuality (Korte et al., 2008). Research in desistance confirms that the majority of youth whose gender dysphoria resolves naturally do indeed grow up to be gay, lesbian, or bisexual adults (Cantor, 2020, Appendix; Singh et al., 2021).

ii. Implications of very low-quality evidence that underlies the practice of pediatric gender transition are not explained

The quality of evidence underlying the practice of pediatric gender transition is widely recognized to be of very low quality (Hembree et al., 2017). In 2020, the most comprehensive systematic review of evidence to date, commissioned by the UK National Health System (NHS) and conducted by the National Institute for Health and Care Excellence (NICE), concluded that the evidence for both puberty blocking and cross-sex hormones is of very low certainty (National Institute for Health & Care Excellence, 2020a; 2020b).

According to the NICE review of evidence for puberty blockers, the studies "are all small, uncontrolled observational studies, which are subject to bias and confounding, and are of very low certainty as assessed using modified GRADE [Grading of Recommendations, Assessment, Development and Evaluations]. All the included studies reported physical and mental health comorbidities and concomitant treatments very poorly" (National Institute for Health & Care Excellence, 2020a, p.13). NICE reached similar conclusions regarding the quality of the evidence for cross-sex hormones (National Institute for Health & Care Excellence, 2020b).

Problematically, the implications of administering a treatment with irreversible, life-changing consequences based on evidence that has an official designation of "very low certainty" according to modified GRADE is rarely discussed with the patients and the families. GRADE is the most widely adopted tool for grading the quality of evidence and for making treatment recommendations worldwide. GRADE has four levels of evidence, also known as certainty in evidence or quality of evidence: very low, low, moderate, and high (BMJ Best Practice, 2021). When evidence is assessed to be "very low certainty," there is a high likelihood that the patients will not experience the effects of the proposed interventions (Balshem et al., 2011).

In the context of providing puberty blockers and cross-sex hormones, the designation of "very low certainty" signals that the body of evidence asserting the benefits of these interventions is

highly unreliable. In contrast, several negative effects are quite certain. For example, puberty blockade followed by cross-sex hormones leads to infertility and sterility (Laidlaw, Van Meter, Hruz, Van Mol, & Malone, 2019). Surgeries to remove breasts or sex organs are irreversible. Other health risks, including risks to bone and cardiovascular health, are not fully understood and are uncertain, but the emerging evidence is alarming (Alzahrani et al., 2019; Biggs, 2021).

iii. The question of suicide is inappropriately handled

Suicide among trans-identified youth is significantly elevated compared to the general population of youth (Biggs, 2022; de Graaf et al., 2020). However, the "transition or die" narrative, whereby parents are told that their only choice is between a "live trans daughter or a dead son" (or vice-versa), is both factually inaccurate and ethically fraught. Disseminating such alarmist messages hurts the majority of trans-identified youth who are not at risk for suicide. It also hurts the minority who are at risk, and who, as a result of such misinformation, may forgo evidence-based suicide prevention intervention in the false hopes that transition will prevent suicide.

The notion that trans-identified youth are at alarmingly high risk of suicide usually stems from biased online samples that rely on self-report (D'Angelo et al., 2020; James et al., 2016; The Trevor Project, 2021), and frequently conflates suicidal thoughts and non-suicidal self-harm with serious suicide attempts and completed suicides. Until recently, little was known about the actual rate of suicide of trans-identified youth. However, a recent analysis of data from the biggest pediatric gender clinic in the world, the UK's Tavistock, found the rate of completed youth suicides to be 0.03% over a 10-year period, which translates into the annual rate of 13 per 100,000 (Biggs, 2022). While this rate is significantly elevated compared to the general population of teens, it is far from the epidemic of trans suicides portrayed by the media.

The "transition or die" narrative regards suicidal risk in trans-identified youth as a different phenomenon than suicidal risk among other youth. Making them an exception falsely promises the parents that immediate transition will remove the risk of suicidal self-harm. Trans patients themselves complain about the so-called "trans broken arm syndrome" – a frustrating pattern whereby physicians "blame" all the problems the patients are experiencing on their trans status, and a result, fail to perceive and respond to other sources of distress (Paine, 2021). Clinicians caring for trans-identified youth should be reminded that suicide risk in all patients is a multi-factorial phenomenon (Mars et al., 2019). To treat trans youths' suicidality as an exception is to deny them evidence-based care.

A recent study of three major youth clinics concluded that suicidality of trans-identifying teens is only somewhat elevated compared to that of youth referred for mental health issues unrelated to gender identity struggles (de Graaf et al., 2020). Another study found that transgender-identifying teens have relatively similar rates of suicidality compared to teens who are gay, lesbian and bisexual (Toomey, Syvertsen, & Shramko, 2018). Depression, eating disorders, autism spectrum conditions, and other mental health conditions commonly found in transgender-identifying youth (Kaltiala-Heino, Bergman, Työläjärvi, & Frisen, 2018; Kozlowska, McClure, et al., 2021; Morandini, Kelly, de Graaf, Carmichael, & Dar-Nimrod, 2021) are all known to independently contribute to the probability of suicide (Biggs, 2022; Simon & VonKorff, 1998; Smith, Zuromski, & Dodd, 2018).

The "transition or suicide" narrative falsely implies that transition will prevent suicides. Clinicians working with trans-identified youth should be aware that although in the short-term, gender-affirmative interventions can lead to improvements in some measures of suicidality (Kaltiala et al., 2020), neither hormones nor surgeries have been showed to reduce suicidality in the long-term (Bränström & Pachankis, 2020a; 2020b). Alarmingly, a longitudinal study from Sweden that covered more than a 30-year span found that adults who underwent surgical transition were 19 times more likely than their age-matched peers to die by suicide overall, with female-to-male participants' risk 40 times the expected rate (Dhejne et al., 2011, Table S1).

Another key longitudinal study from the Netherlands concluded that suicides occur at a similar rate at all stages of transition, from pretreatment assessment to post-transition follow-up (Wiepjes et al., 2020). The data from the Tavistock clinic also did not show a statistically significant difference between completed suicides in the "waitlist" vs. the "treated" groups (Biggs, 2022). Luckily, in both groups, completed suicides were rare events (which may have been responsible for the lack of statistical significance). Thus, we consider the "transition or die" narrative to be misinformed and ethically wrong.

In our experience in working with trans-identified youth, an adolescent's suicidality can sometimes arise as a response to parental distress, resistance, skepticism, or wish to investigate the forces shaping the new gender identity before social transition and hormone therapy. When mental health professionals or other healthcare providers fail to recognize the legitimacy of parental concerns, or label the parents as transphobic, this only tends to intensify intrafamilial tension. Clinicians would be well-advised that gender transition is not an appropriate response to suicidal intent or threat, as it ignores the larger mental health and social context of the young patient's life—the entire family is often in crisis. Trans-identified adolescents should be screened for self-harm and suicidality, and if suicidal behaviors are present, an appropriate evidence-based suicide prevention plan should be put in place (de Graaf et al., 2020).

The Dutch Study: the questionable basis for the gender affirmative model of care for youth

Few practitioners of gender-affirmative interventions, and even fewer patients and families, realize that the foundation of the practice of medically transitioning minors stems from a single Dutch proof of concept study, the outcomes of which were documented in two studies (de Vries, Steensma, Doreleijers, Cohen, & Kettenis, 2011; de Vries et al., 2014). The former (de Vries et al., 2011) reported on cases who underwent puberty blockade, while the latter (de Vries et al., 2014) reported on a subset of the cases who completed surgeries.

The Dutch study subjects' high level of psychological functioning at 1.5 years after surgery, which was the study end point, was an impressive feat. However, both of the studies suffer from a high risk of bias due to their study design, which is effectively a non-randomized case series—one of the lowest levels of evidence (Mathes & Pieper, 2017; National Institute for Health & Care Excellence, 2020a). In addition, the studies suffer from limited applicability to the populations of adolescents presenting today (de Vries, 2020). The interventions described in the study are currently being applied to adolescents who were not cross-gender identified prior to puberty, who have significant mental health problems, as well as those who have non-binary identities—all of these presentations were explicitly disqualified from the Dutch protocol. Despite these limitations, the Dutch clinical experiment has become the basis for the practice of medical transition of minors worldwide and serves as the basis for the recommendations outlined in the 2017 Endocrine Society guidelines (Hembree et al., 2017).

We contend that the Dutch studies have been misunderstood and misrepresented as providing evidence of the safety and efficacy of these interventions for all youth. It is important that both the strengths and the weaknesses of these two studies are understood, as to date, the Dutch experience presents the best available evidence behind the practice of pediatric gender transition.

Rationale for pediatric transition

Prior to the 1990s, gender transitions were typically initiated in mature adults (Dhejne et al., 2011). However, it was noted that particularly for natal male patients, hormonal and surgical interventions failed to achieve satisfactory results, and patients had a "never disappearing masculine appearance" (Delemarre-van de Waal & Cohen-Kettenis, 2006). The lack of adequate cosmetic outcomes was thought to contribute to the frequently disappointing outcomes of medical

gender transition, with persistently high rates of mental illness and suicidality post-transition (Delemarre-van de Waal & Cohen-Kettenis, 2006; Dhejne et al., 2011; Ross & Need, 1989).

In the mid 1990s, a team of Dutch researchers hypothesized that by carefully selecting a subset of gender dysphoric children who would likely be transgender-identified for the rest of their lives, and by medically intervening before puberty left an irreversible mark on their bodies, the cosmetic outcomes would be improved—and as a result, mental health outcomes might be improved (Gooren & Delemarre-van de Waal, 1996).

Mixed study findings

In 2014, the Dutch research team published a key longitudinal study of mental health outcomes of 55 youths who completed medical and surgical transition (de Vries et al., 2014). The 2014 paper (sometimes referred to as the "Dutch study") reported that for youth with severe gender dysphoria that started in early childhood and persisted into mid-adolescence, a sequence of puberty blockers, cross-sex hormones, and breast and genital surgeries (including a mandatory removal of the ovaries, uterus and testes), with ongoing extensive psychological support, was associated with positive mental health and overall function 1.5 years post-surgery.

While the Dutch reported resolution of gender dysphoria post-surgery in study subjects, the reported psychological improvements were quite modest (de Vries et al., 2014). Of the 30 psychological measurements reported, nearly half showed no statistically significant improvements, while the changes in the other half were marginally clinically significant at best (Malone, D'Angelo, et al., 2021). The scores in anxiety, depression, and anger did not improve. The change in the Children's Global Assessment Scale, which measures overall function, was one of the most impressive changes—however it too remained in the same range before and after treatment (de Vries et al., 2014).

Problematic discordance between reduced gender dysphoria and lack of meaningful improvements in psychological measures

The discordance between the marked reduction in gender dysphoria, as measured by the UGDS (Utrecht Gender Dysphoria Scale), and the lack of meaningful changes in psychological function using standard measures, warrants further examination. There are three plausible explanations for this lack of agreement. Any one of these three explanations calls into question the widely assumed notion that the medical interventions significantly improve mental health or lessen or eradicate gender dysphoria.

One possible explanation is that gender dysphoria as measured by UGDS, and psychological function, as measured by most standard instruments, are not correlated. This contradicts the primary rationale for providing gender-affirmative treatments for youth (which is to improve psychological health and functioning), and if true, ethically threatens these medical interventions. The other plausible explanation stems from the high psychological function of all the subjects at baseline; the subjects were selected because they were free from significant mental health problems (de Vries et al., 2014). As a result, there was little opportunity to meaningfully improve. This explanation highlights a key limitation in applying the study's results to the majority of today's gender dysphoric youth, who often present with a high burden of mental illness (Becerra-Culqui et al., 2018; Kozlowska, McClure, et al., 2021). The study cannot be used as evidence that these procedures have been proven to improve depression, anxiety, and suicidality.

A third possible explanation for the discordance between only minor changes in psychological outcomes but a significant drop in gender dysphoria comes from a close examination of the UGDS scale itself and how it was used by the Dutch researchers. This 12-item scale, designed by the Dutch to assess the severity of gender dysphoria and to identify candidates for hormones

and surgeries, consists of "male" (UGDS-aM) and "female" (UGDS-aF) versions (Iliadis et al., 2020). At baseline and after puberty suppression, biological females were given the "female" scale, while males were given the "male" scale. However, post-surgery, the scales were flipped: biological females were assessed using the "male" scale, while biological males were assessed on the "female" scale (de Vries et al., 2014). We maintain that this handling of the scales may have at best obscured, and at worst, severely compromised the ability to meaningfully track how gender dysphoria was affected throughout the treatment.

Consider this example. At baseline, a gender dysphoric biological female would rate items from the "female" scale such as: "I prefer to behave like a boy" (item 1); "I feel unhappy because I have to behave like a girl" (item 6) and "I wish I had been born a boy" (item 12). Positive answers to these questions would have contributed to a high baseline gender dysphoria score. After the final surgery, however, this same patient would be asked to rate items from the "male" scale, including the following: "My life would be meaningless if I had to live as a boy" (item 1); "I hate myself because I am a boy" (item 6) and "It would be better not to live than to live as a boy" (item 12). A gender dysphoric female would not endorse these statements (at any stage of the intervention), which would lead to a lower gender dysphoria score.

Thus, the detected drop in the gender dysphoria scores for biological males and females may have had less to do with the success of the interventions, and more to do with switching the scale from the "female" to the "male" version (and vice-versa) between the baseline and post-surgical period. This, too, may explain why no changes in gender dysphoria were noted between baseline and the puberty blockade phase, and were only recorded after the final surgery, when the scale was switched.

It must be considered that had the researchers administered the "flipped" scale earlier, at the completion of the puberty blocker stage, UGDS scale could have registered the reduction in gender dysphoria. Likewise, however, one must consider the possibility that had *both sets of scales* been administered to the same individual at baseline, a "reduction" in gender dysphoria could have been registered upon switching of the scale, *well before any interventions began*. The question here is whether the diminishment of quantitative measures of gender dysphoria is largely an artifact of what scale was used.

It must be noted that the UGDS measure has been demonstrated only to effectively differentiate between clinically referred gender dysphoric individuals, non-clinically referred controls, and participants with disorders of sexual development, and was not designed to detect changes in gender dysphoria during treatment (Steensma, McGuire, Kreukels, et al. 2013). The presence of items such as "I dislike having erections" (item 11, UGDS-aM), which would have to be rated by birth-females, and "I hate menstruating because it makes me feel like a girl (item 10, UGDS-aF), which would be presented to birth-males, neither of which could be meaningfully rated by either at any stage of the interventions, further illustrates that UGDS has questionable validity for the purpose of detecting meaningful changes in gender dysphoria as a result of medical and surgical treatment.

The updated UGDS scale (UGDS-GS), developed by the Dutch after the publication of their seminal study, has eliminated the two-sex version of the scale in favor of a single battery of questions applicable to both sexes (McGuire et al., 2020). This change may lead to a more reliable measurement of treatment-associated changes in future research. Other gender dysphoria scales also exist (Hakeem, Črnčec, Asghari-Fard, Harte, & Eapen, 2016; Iliadis et al., 2020) and may or may not be better suited for the purposes of measuring the impact of medical interventions on underlying gender distress. Gender dysphoria, of course, may also prove to be a more complex concept than can be measured by any scale.

Other limitations

The two Dutch studies were conducted without a control group (de Vries et al., 2011; de Vires et al., 2014). Nor could the researchers control for mental health interventions, which all the

subjects received in addition to hormones and surgery. The Dutch only evaluated mental health outcomes and did not assess physical health effects of hormones and surgery. The sample size was small: the final study reported the outcomes of only 55 children, and as few as 32 were evaluated on key measures of psychological outcomes.

It is important to realize that the Dutch sample was carefully selected, which introduced a source of bias, and also challenges the study's applicability. From the 196 adolescents initially referred, 111 were considered eligible to start puberty blockers, and of this group, only the 70 most mature and mentally stable who proceeded to cross-sex hormones were included in the study (de Vries et al., 2011). Of note, 97% of the selected cases were attracted to members of their natal sex at baseline. All were cross-sex identified, with no cases of non-binary identities. The final study only followed 55, rather than the original 70 cases, further excluding from reporting the outcomes of subjects who had experienced adverse events, including: one death from surgery-related complications and three cases of complications such as obesity and diabetes that rendered subjects ineligible for surgery. Three more subjects refused to be contacted or dropped out of care, which may mask adverse outcomes (de Vries et al., 2014).

There is no knowledge of the fate of 126 patients who did not participate in the Dutch study. Longer term outcomes of the subjects who did participate are lacking. We are aware of only one case of long-term follow-up for a female-to-male patient treated by the Dutch team in the 1990s. The case study describing the subject's functioning at the age of 33 found that the patient did not regret gender transition. However, he reported struggling with significant shame related to the appearance of his genitals and to his inability to sexually function; had problems maintaining long-term relationships; and experienced depressive symptoms (Cohen-Kettenis, Schagen, Steensma, de Vries, & Delemarre-van de Waal, 2011). Notably, these problems had not yet emerged when the same patient was assessed at the age of 20, when he reported high levels of satisfaction in general, and was "very satisfied with the results [of the metoidioplasty]" in particular (Cohen-Kettenis & van Goozen, 1998, p.248). Since the last round of psychological outcomes of the individuals in the Dutch study was obtained when the subjects were around 21 years of age (de Vries et al., 2014), it raises questions how they will fair in during the decade when new developmental tasks, such as, career development, forming long-term intimate relationships and friendships, or starting families come into focus.

As to the unknown outcomes of the patients rejected by the Dutch protocol, one study did report on 14 adolescents who sought gender reassignment in the same clinic, but were disqualified from treatment due to "psychological or environmental problems" (Smith, Van Goozen, & Cohen-Kettenis, 2001, p. 473). The study found that at follow-up 1-7 years after the original application, 11 of the 14 no longer wished to transition, and 2 others only slightly regretted not transitioning (Malone, D'Angelo, et al., 2021; Smith et al., 2001). This further underscores the importance of conducting research utilizing control groups and following the subjects for an extended period.

A recent attempt to replicate the results of the first Dutch study (de Vries et al., 2011) found no demonstrable psychological benefit from puberty blockade, but did find that the treatment adversely affected bone development (Carmichael et al., 2021). The final Dutch study (de Vries et al., 2014) has never been attempted to be replicated with or without a control group.

The scaling of the Dutch Protocol beyond original indications

The medical and surgical sequence of Dutch protocol has been aggressively scaled worldwide without the careful evaluations and vetting practiced by the Dutch. The protocol's original investigators have recently expressed concern that the interventions they described have been widely adopted on four continents without several of the protocol's essential discriminatory features (de Vries, 2020).

The extensive multi-year multidisciplinary evaluations of the children have been abbreviated or simply bypassed. The medical sequence is routinely used for children with post-pubertal onset of transgender identities complicated by mental health comorbidities (Kaltiala-Heino et al., 2018), and not just for those high-functioning adolescents with persistent early life cross-identifications, as was required by the Dutch protocol (de Vries & Cohen-Kettenis, 2012). Further, it has become increasingly common to socially transition children before puberty (Olson, Durwood, DeMeules, & McLaughlin, 2016), even though this was explicitly discouraged by the Dutch protocol at the time (de Vries & Cohen-Kettenis, 2012).

In addition, medical transition is frequently initiated much earlier than recommended by the original protocol (de Vries & Cohen-Kettenis, 2012). The authors of the protocol were aware that most children would have a spontaneous realignment of their gender identity with sex by going through early- to mid-stages of puberty (Cohen-Kettenis, Delemarre-van de Waal, & Gooren, 2008). The average age of initiating puberty blockade in the Dutch study was around 15. In contrast, currently the age limit has been lowered to the age of Tanner stage II, which can occur as early as 8-9 years (Hembree et al., 2017). Irreversible cross-sex hormones, initiated in the Dutch study at the average age of nearly 17, are currently commonly prescribed to 14-year-olds, and this lower age threshold has been recommended by draft recommendation by WPATH Standards of Care 8, the final version of which is due to be released in early 2022. The fact that children are transitioned before their identity is tested against the biological reality and before natural resolution of gender dysphoria has had a chance to occur is a major deviation from the original Dutch protocol. Systematic follow-up, reassessments, and tracking and publishing of outcomes are not performed.

As the lead Dutch researchers have begun to call for more research into the novel presentation of gender dysphoria in youth (de Vries, 2020; Voorzij, 2021) and question the wisdom of applying the hormonal and surgical treatment protocols to the newly presenting cases, many recently educated gender specialists mistakenly believe that the Dutch protocol proved the concept that its sequence helps all gender-dysphoric youth. Although aware of the Dutch study's importance, they seem to be unaware of its agreed upon limitations, and the Dutch clinicians' own discomfort that most new trans-identified adolescents presenting for care today significantly differ from the population the Dutch had originally studied. These facts, of course, underscore the need for a robust informed consent process.

The recommendations for informed consent process for children, adolescents, and young adults

Consent for all stages of gender transition should be explicit, not implied

Noninvasive medical care or care that carries little risk of harm does not require a signed informed consent document; rather, consent is implied through the act of a patient presenting for care. For example, when a parent brings in a child for a skin laceration or abscess, consent for sutures or simple incision and drainage is implied. Similarly, when a child presents with pneumonia and is hospitalized, consent for chest x-ray, IV fluids, and antibiotics is also implied. It is assumed that patients or their guardians agree to the interventions and understand the benefits and risks. When risks are greater, such as prior to surgery, chemotherapy, or another invasive procedure, an informed consent document is signed. Such situations require an explicit, or express informed consent.

In the context of interventions for gender dysphoria or gender incongruence, the uncertainties associated with puberty blocking, cross-sex hormones, and gender-affirmative surgeries are well-recognized (Manrique et al., 2018; National Institute for Health & Care Excellence, 2020a; 2020b; Wilson et al., 2018). In these cases, consent should be explicit rather than implied because of the complexity, uncertainty, and risks involved.

Informed consent for social transition represents a gray area. Evidence suggests that social transition is associated with the persistence of gender dysphoria (Hembree et al.,

2017; Steensma, McGuire, Kreukels, Beekman, & Cohen-Kettenis, 2013). This suggests that social gender transition is a form of a psychological intervention with potential lasting effects (Zucker, 2020). While the causality has not been proven, the possibility of iatrogenesis and the resulting exposure to the risks of future medical and surgical gender dysphoria treatments, qualifies social gender transition for explicit, rather than implied, consent.

Full unbiased disclosure of benefits, risks and alternatives is requisite

When mental health professionals are involved in evaluations and recommendations, the informed consent process begins either as part of an extended evaluation or is integrated in a psychotherapeutic process, separately or together, with the parents and patient. When pediatricians, nurse practitioners, or primary care physicians perform the initial evaluation, the informed consent process is more likely to be labeled as such in a briefer series of meetings.

In all settings, the informed consent discussions for gender-affirmative care should include three central ideas:

1. The decision to initiate gender transition may predispose the child to persist in their transgender identity long-term.
2. Many of the physical changes contemplated and undertaken are irreversible.
3. Careful long-term studies have not been done to verify that these interventions enable better physical and mental health or improved social functioning, or that they do not cause harm.

The informed consent process, culminating with a signed document, signifies that parents and patient have been educated about the short- and long-term risks, benefits and uncertainties associated with all relevant stages of the gender-affirmative interventions. The process must also inform the patients and families about the full range of alternative treatments, including the choice of not socially or medically treating the child's or adolescent's current state of gender/body incongruence.

Decisional capacity to consent needs to be assessed and family should be involved

Trans-identified youth typically present themselves as strongly desiring hormones and ultimately, surgery. It should not be assumed that their eagerness is matched with the capacity to carefully consider the consequences of their realized desires. Trans-identified youth younger than the age of consent should be part of the informed consent process, but they may not be mature enough to recognize or admit their concerns about the proposed intervention. For this reason, it is the parents who, after careful consideration, are responsible for signing an informed consent document.

The issue of the exact age at which adolescents are mature enough to consent to gender transition has proven contentious: courts have been asked to decide about competence to consent to gender-affirmative hormones for youth in the United Kingdom and Australia (Ouliaris, 2021). In the United States, the legal age for medical consent for gender-affirmative interventions varies by state.

When patients are age 18 and older, and in some jurisdictions as young as age 15 (Right to medical or dental treatment without parental consent, 2010), they do not legally require parental approval for medical procedures. But because an individual's change of gender has profound implications for parents, siblings, and other family members, it is usually prudent for clinicians to seek their input directly or indirectly during the informed consent process. This is done by requesting a meeting with the parents.

A recent study by a Dutch research team attempted to evaluate the decisional capacity of adolescents embarking on gender transition (Vrouenraets, de Vries, de Vries, van der Miesen, & Hein, 2021). The researchers administered the MacCAT-T tool, comprised of the *understanding*, *appreciating*, *reasoning*, and *expressing a choice* domains, to 74 adolescents who were 14.7 years old on average (with the minimum age of 10). They concluded that the adolescents were competent to consent for starting pubertal suppression, calling for similar research for the <12 group, particularly because "birth-assigned girls … may benefit from puberty suppression as early as 9 years of age" (Vrouenraets et al., 2021 p.7).

This study suffers from two significant limitations involving the MacCAT-T tool. It was never designed for children. Rather, it was designed to assess medical consent capacities of adults suffering from conditions such as dementia, schizophrenia, and other psychiatric disorders. There is a fundamental lack of equivalency between consenting to treatment by adults with cognitive impairments and obtaining consent from healthy children whose age-appropriate cognitive capacities are intact, but who lack the requisite life experiences to consent to profound life-changing medical interventions. We doubt, for example, whether even highly intelligent children who have not had sexual experiences can meaningfully comprehend the loss of future sexual function and reproductive abilities.

In addition, even for adults, the MacCAT-T tool has been criticized for its exclusive focus on cognitive aspects of capacity, failing to account for the non-cognitive aspects such as values, emotions and other biographic and context specific aspects inherent in the complexity of the decision process in real life (Breden & Vollmann, 2004). Children's values and emotions undergo tremendous change during the process of maturation.

The authors' conclusion about their young patients' competence to consent should be compared with what a panel of judges wrote in the challenge to the Tavistock treatment protocol (Bell v Tavistock, 2020):

> …the clinical intervention we are concerned with here is different in kind to other treatments or clinical interventions. In other cases, medical treatment is used to remedy, or alleviate the symptoms of, a diagnosed physical or mental condition, and the effects of that treatment are direct and usually apparent. The position in relation to puberty blockers would not seem to reflect that description. [para 135]

> …we consider the treatment in this case to be in entirely different territory from the type of medical treatment which is normally being considered. [para 140]

> … the combination here of lifelong and life changing treatment being given to children, with very limited knowledge of the degree to which it will or will not benefit them, is one that gives significant grounds for concern. [para 143]

It seems clear that perceptions of children as young as 10 years of age as medically competent vary by country, state, and the institution where the doctor works, and, by clinicians' beliefs about the long-term benefits of these interventions. We maintain that the claim that kids can consent to extreme life-altering interventions is a fundamentally a philosophical claim (Clark & Virani, 2021). Our view in this matter is that consent is primarily a parental function.

Informed consent should be viewed as a process rather than an event

Most institutions that care for transgender-identified individuals have devised obligatory consent forms that outline the risks and uncertainties of hormonal and surgical gender-affirmative interventions. However, the requisite signatures are frequently collected in a perfunctory manner (Schulz, 2018), akin to signatures collected ahead of a common surgical procedure. The purpose of such informed consent documents appears to be to protect practitioners from lawsuits, rather than attend to the primary ethical foundation of the process.

Although obtaining the signatures is important, the signed document should signify that the process of informed consent has been undertaken over an extended time period and is not simply quickly completed (Vrouenraets et al., 2021). We believe the latter approach poses an ethical concern (Levine, 2019).

The internal dynamics of the trans-identified young person and their families vary considerably. Parental capacities, their private marital and intrafamilial relationships, their cultural awareness, religious and political sensibilities all influence the amount of time necessary to undertake a thorough informed consent process. It is not prudent to suggest a specific duration for the process of informed consent, other than to emphasize that it requires a slow, patient, thoughtful question and answer period as the parents and patient contemplate the meaning of what is known and unknown and whether to embark on alternative approaches to the management of gender dysphoria before the age of full neurological maturity has been reached, mental health comorbidities have been addressed, and a true informed consent by the patient is more likely.

Final thoughts

Sixty years of experience providing medical and surgical assistance to transgender-identified persons have seen many changes in who is treated, when they are treated, and how they are treated. Today, the emphasis has shifted to the treatment of the unprecedented numbers of youth declaring a trans identity. As adolescents pursue social, medical, and surgical interventions, health care providers may experience unease about patients' cognitive and emotional capacities to make decisions with life-changing and enduring consequences. An unrushed informed consent process helps the provider, the parents, and the patient.

Three issues tend to obscure the salience of informed consent: conspicuous mental health problems, uncertainty about the minor's personal capacity to understand the irreversible nature of the interventions, and parental disagreement. Physical and psychiatric comorbidities can contribute to the formation of a new identity, develop as its consequence, or bear no connection to it. Assessing mental health and the minor's functionality is one of the reasons why rapid affirmative care may be dangerous for patients and their families. For example, when situations involve autism, learning disorders, sexual abuse, attachment problems, trauma, separation anxiety, previous depressed or anxious states, neglect, low IQ, past psychotic illness, eating disorders or parental mental illness, clinicians must choose between ignoring these potentially causative conditions and comorbidities and providing appropriate treatment before affirmative care (D'Angelo et al., 2020).

For youth less than the age of majority, informed consent via parents provides a legal route for treatment but it does not make the decisions to transition, provide hormones, or surgically remove breasts or testes less fraught with uncertainty. The best that health professionals can do is to ensure that the consent process informs the patient and parents of the current state of science, which is sorely lacking in quality research. It is the professionals' responsibility to ensure that the benefits patients and parents seek, and the risks they are assuming, are clearly appreciated as they prepare to make this often-excruciating decision.

Young people who have reached the age of majority, but who have not reached full maturation of the brain represent a unique challenge. It is well-recognized that brain remodeling proceeds through the third decade of life, with the prefrontal cortex responsible for executive function and impulse control the last to mature (Katz et al., 2016). The growing number of detransitioners who had been old enough to legally consent to transition, but who no longer felt they were transgender upon reaching their mid-20's, raises additional concerns about this vulnerable age group (Littman, 2021; Vandenbussche, 2021).

When the clinician is uncertain whether a young person is competent to comprehend the implications of the desired treatment—that is, when informed consent cannot inform the patient—the clinician may need more time with the patient. When parents or guardians do

not agree about whether to use puberty blockers or cross-sex hormones, clinicians are in an uneasy spot (Levine, 2021). This occurs in both intact and divorced families. Australia has given legal instructions to clinicians facing these uncertainties: the court is to be asked to decide (Ouliaris, 2021). The court system in the UK has been grappling with similar issues in recent years. While it is a rare case that ends up in a courtroom, clinicians devoted to a deliberate informed consent process are still likely to encounter ethical dilemmas that they cannot resolve.

Acknowledgments

The authors wish to thank SEGM staff for their grant and bibliographic support.

Funding

This work was supported by the Society for Evidence based Gender Medicine.

References

Aitken, M., Steensma, T. D., Blanchard, R., VanderLaan, D. P., Wood, H., Fuentes, A., Spegg, C., Wasserman, L., Ames, M., Fitzsimmons, C. L., Leef, J. H., Lishak, V., Reim, E., Takagi, A., Vinik, J., Wreford, J., Cohen-Kettenis, P. T., de Vries, A. L. C., Kreukels, B. P. C., & Zucker, K. J. (2015). Evidence for an altered sex ratio in clinic-referred adolescents with gender dysphoria. *The Journal of Sexual Medicine*, *12*(3), 756–763. doi:10.1111/jsm.12817

Alzahrani, T., Nguyen, T., Ryan, A., Dwairy, A., McCaffrey, J., Yunus, R., Forgione, J., Krepp, J., Nagy, C., Mazhari, R., & Reiner, J. (2019). Cardiovascular disease risk factors and myocardial infarction in the transgender population. *Circulation: Cardiovascular Quality and Outcomes*, *12*(4). doi:10.1161/CIRCOUTCOMES.119.005597

American College Health Association. (2021). *American College Health Association-National College Health Assessment III: Undergraduate Student Reference Group Data Report Spring 2021*. Boston: ACHA-NCHA III. https://www.acha.org/documents/ncha/NCHA-III_SPRING-2021_UNDERGRADUATE_REFERENCE_GROUP_DATA_REPORT.pdf

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). doi:10.1176/appi.books.9780890425596

Anderson, E. (2022, January 3). Opinion: When it comes to trans youth, we're in danger of losing our way. *The San Francisco Examiner.* Retrieved January 5, 2022, from http://www.sfexaminer.com/opinion/are-we-seeing-a-phenomenon-of-trans-youth-social-contagion/

Anderson, E., Edwards-Leeper, L. (2021, November 24). The mental health establishment is failing trans kids. Washington, DC: Washington Post. Retrieved December 20, 2021, from https://www.washingtonpost.com/outlook/2021/11/24/trans-kids-therapy-psychologist/

Asscheman, H., Giltay, E. J., Megens, J. A. J., de Ronde, W. (Pim), van Trotsenburg, M. A. A., & Gooren, L. J. G. (2011). A long-term follow-up study of mortality in transsexuals receiving treatment with cross-sex hormones. *European Journal of Endocrinology*, *164*(4), 635–642. doi:10.1530/EJE-10-1038

Balshem, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., Vist, G. E., Falck-Ytter, Y., Meerpohl, J., & Norris, S. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, *64*(4), 401–406. doi:10.1016/j.jclinepi.2010.07.015

Becerra-Culqui, T. A., Liu, Y., Nash, R., Cromwell, L., Flanders, W. D., Getahun, D., Giammattei, S. V., Hunkeler, E. M., Lash, T. L., Millman, A., Quinn, V. P., Robinson, B., Roblin, D., Sandberg, D. E., Silverberg, M. J., Tangpricha, V., & Goodman, M. (2018). Mental health of transgender and gender nonconforming youth compared with their peers. *Pediatrics*, *141*(5), e20173845. doi:10.1542/peds.2017-3845

Bechard, M., VanderLaan, D. P., Wood, H., Wasserman, L., & Zucker, K. J. (2017). Psychosocial and psychological vulnerability in adolescents with gender dysphoria: A "proof of principle" Study. *Journal of Sex & Marital Therapy*, *43*(7), 678–688. doi:10.1080/0092623X.2016.1232325

Bell v Tavistock and Portman NHS Foundation Trust. (2020). EWHC 3274. The High Court of Justice (2020). https://www.judiciary.uk/wp-content/uploads/2020/12/Bell-v-Tavistock-Judgment.pdf

Biggs, M. (2021). Revisiting the effect of GnRH analogue treatment on bone mineral density in young adolescents with gender dysphoria. *Journal of Pediatric Endocrinology and Metabolism*. doi:10.1515/jpem-2021-0180

Biggs, M. (2022). Suicide by clinic-referred transgender adolescents in the United Kingdom. *Archives of Sexual Behavior*.

BMJ Best Practice. (2021). What is grade? Retrieved January 1, 2022, from https://bestpractice.bmj.com/info/us/toolkit/learn-ebm/what-is-grade/

Bonfatto, M., & Crasnow, E. (2018). Gender/ed identities: An overview of our current work as child psychotherapists in the Gender Identity Development Service. *Journal of Child Psychotherapy*, *44*(1), 29–46. doi:10.1080/0075417X.2018.1443150

Boyd, I., Hackett, T., & Bewley, S. (2022). Care of transgender patients: A general practice quality improvement approach. *Healthcare*, *10*(1), 121. doi:10.3390/healthcare10010121

Bränström, R., & Pachankis, J. E. (2020a). Reduction in mental health treatment utilization among transgender individuals after gender-affirming surgeries: A total population study. *American Journal of Psychiatry*, *177*(8), 727–734. doi:10.1176/appi.ajp.2019.19010080

Bränström, R., & Pachankis, J. E. (2020b). Correction to Bränström and Pachankis. (2020). *American Journal of Psychiatry*, *177*(8), 734–734. doi:10.1176/appi.ajp.2020.1778correction

Breden, T. M., & Vollmann, J. (2004). The Cognitive Based Approach of Capacity Assessment in Psychiatry: A Philosophical Critique of the MacCAT-T. *Health Care Analysis*, *12*(4), 273–283. doi:10.1007/s10728-004-6635-x

Bridge, J. A., Greenhouse, J. B., Ruch, D., Stevens, J., Ackerman, J., Sheftall, A. H., Horowitz, L. M., Kelleher, K. J., & Campo, J. V. (2020). Association Between the Release of Netflix's 13 Reasons Why and Suicide Rates in the United States: An Interrupted Time Series Analysis. J Am Acad Child Adolesc Psychiatry, *59*(2), 236–243. doi:10.1016/j.jaac.2019.04.020

Byne, W., Bradley, S.J., Coleman, E., Eyler, A.E., Green, R., Menvielle, E.J., Meyer-Bahlburg, H.F.L., Pleak, R.R. & Tompkins, D.A. (2012). Report of the American Psychiatric Association Task Force on Treatment of Gender Identity Disorder. Archives of Sexual Behavior, *41*(4):759–796. doi:10.1007/s10508-012-9975-x

Cantor, J. M. (2020). Transgender and gender diverse children and adolescents: Fact-checking of AAP Policy. *Journal of Sex & Marital Therapy*, *46*(4), 307–313. doi:10.1080/0092623X.2019.1698481

Carmichael, P., Butler, G., Masic, U., Cole, T. J., De Stavola, B. L., Davidson, S., Skageberg, E. M., Khadr, S., & Viner, R. M. (2021). Short-term outcomes of pubertal suppression in a selected cohort of 12 to 15 year old young people with persistent gender dysphoria in the UK. *PLOS ONE*, *16*(2), e0243894. doi:10.1371/journal.pone.0243894

Chew, D., Tollit, M. A., Poulakis, Z., Zwickl, S., Cheung, A. S., & Pang, K. C. (2020). Youths with a non-binary gender identity: A review of their sociodemographic and clinical profile. *The Lancet Child & Adolescent Health*, *4*(4), 322–330. doi:10.1016/S2352-4642(19)30403-1

Churcher Clarke, A., & Spiliadis, A. (2019). 'Taking the lid off the box': The value of extended clinical assessment for adolescents presenting with gender identity difficulties. *Clinical Child Psychology and Psychiatry*, *24*(2), 338–352. doi:10.1177/1359104518825288

Clark, B. A., & Virani, A. (2021). This Wasn't a "split-second decision": An empirical ethical analysis of transgender youth capacity, rights, and authority to consent to hormone therapy. *Journal of Bioethical Inquiry*, *18*(1), 151–164. doi:10.1007/s11673-020-10086-9

Clayton, A., Malone, W. J., Clarke, P., Mason, J., & D'Angelo, R. (2021). Commentary: The signal and the noise—questioning the benefits of puberty blockers for youth with gender dysphoria—a commentary on Rew et al. (2021). *Child and Adolescent Mental Health*, *27*, camh.12533. doi:10.1111/camh.12533

Cohen-Kettenis, P. T., Delemarre-van de Waal, H. A., & Gooren, L. J. G. (2008). The treatment of adolescent transsexuals: Changing insights. *The Journal of Sexual Medicine*, *5*(8), 1892–1897. doi:10.1111/j.1743-6109.2008.00870.x

Cohen-Kettenis, P. T., Schagen, S. E. E., Steensma, T. D., de Vries, A. L. C., & Delemarre-van de Waal, H. A. (2011). Puberty suppression in a gender-dysphoric adolescent: A 22-year follow-up. *Archives of Sexual Behavior*, *40*(4), 843–847. doi:10.1007/s10508-011-9758-9

Cohen-Kettenis, P. T., & van Goozen, S. H. M. (1998). Pubertal delay as an aid in diagnosis and treatment of a transsexual adolescent. *European Child & Adolescent Psychiatry*, *7*(4), 246–248. doi:10.1007/s007870050073

D'Angelo, R. (2018). Psychiatry's ethical involvement in gender-affirming care. *Australasian Psychiatry*, *26*(5), 460–463. doi:10.1177/1039856218775216

D'Angelo, R., Syrulnik, E., Ayad, S., Marchiano, L., Kenny, D. T., & Clarke, P. (2020). One size does not fit all: In support of psychotherapy for gender dysphoria. *Archives of Sexual Behavior*, *50*, 7–16. doi:10.1007/s10508-020-01844-2

de Graaf, N. M., Giovanardi, G., Zitz, C., & Carmichael, P. (2018). Sex ratio in children and adolescents referred to the gender identity development service in the UK (2009–2016)). *Archives of Sexual Behavior*, *47*(5), 1301–1304. doi:10.1007/s10508-018-1204-9

de Graaf, N. M., Steensma, T. D., Carmichael, P., VanderLaan, D. P., Aitken, M., Cohen Kettenis, P. T., de Vries, A., Kreukels, B., Wasserman, L., Wood, H., & Zucker, K. J. (2020). Suicidality in clinic-referred transgender adolescents. *European child & adolescent psychiatry*, *31*, 67–83. doi: 10.1007/s00787-020-01663-9. Advance online publication. doi:10.1007/s00787-020-01663-9

de Vries, A. L. C. (2020). Challenges in timing puberty suppression for gender-nonconforming adolescents. *Pediatrics*, *146*(4), e2020010611. doi:10.1542/peds.2020-010611

de Vries, A. L. C., Beek, T. F., Dhondt, K., de Vet, H. C. W., Cohen-Kettenis, P. T., Steensma, T. D., & Kreukels, B. P. C. (2021). Reliability and clinical utility of gender identity-related diagnoses: comparisons between the ICD-11, ICD-10, DSM-IV, and DSM-5. *LGBT Health*, *8*(2), 133–142. doi:10.1089/lgbt.2020.0272

de Vries, A. L. C., & Cohen-Kettenis, P. T. (2012). Clinical management of gender dysphoria in children and adolescents: The Dutch approach. *Journal of Homosexuality*, *59*(3), 301–320. doi:10.1080/00918369.2012.653300

de Vries, A. L. C., McGuire, J. K., Steensma, T. D., Wagenaar, E. C. F., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2014). Young adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*, *134*(4), 696–704. doi:10.1542/peds.2013-2958

de Vries, A. L. C., Steensma, T. D., Doreleijers, T. A. H., & Cohen-Kettenis, P. T. (2011). Puberty suppression in adolescents with gender identity disorder: A prospective follow-up study. *The Journal of Sexual Medicine*, *8*(8), 2276–2283. doi:10.1111/j.1743-6109.2010.01943.x

Delemarre-van de Waal, H. A., & Cohen-Kettenis, P. T. (2006). Clinical management of gender identity disorder in adolescents: A protocol on psychological and paediatric endocrinology aspects. *European Journal of Endocrinology*, *155*(suppl_1), S131–S137. doi:10.1530/eje.1.02231

Dhejne, C., Lichtenstein, P., Boman, M., Johansson, A. L. V., Långström, N., & Landén, M. (2011). Long-term follow-up of transsexual persons undergoing sex reassignment surgery: Cohort study in Sweden. *PLoS ONE*, *6*(2), e16885. doi:10.1371/journal.pone.0016885

Dhejne, C., Öberg, K., Arver, S., & Landén, M. (2014). An analysis of all applications for sex reassignment surgery in Sweden, 1960–2010: Prevalence, incidence, and regrets. *Archives of Sexual Behavior*, *43*(8), 1535–1545. doi:10.1007/s10508-014-0300-8

Diamond, M. (2013). Transsexuality among twins: Identity concordance, transition, rearing, and orientation. *International Journal of Transgenderism*, *14*(1), 24–38. doi:10.1080/15532739.2013.750222

Ehrensaft, D. (2017). Gender nonconforming youth: Current perspectives. *Adolescent Health, Medicine and Therapeutics*, Volume *8*, 57–67. doi:10.2147/AHMT.S110859

Entwistle, K. (2020). Debate: Reality check – Detransitioner's testimonies require us to rethink gender dysphoria. *Child and Adolescent Mental Health*, *26*, 15–16. camh.12380. doi:10.1111/camh.12380

Frigerio, A., Ballerini, L., & Valdés Hernández, M. (2021). Structural, functional, and metabolic brain differences as a function of gender identity or sexual orientation: A systematic review of the human neuroimaging literature. *Archives of sexual behavior*, *50*(8), 3329–3352. doi:10.1007/s10508-021-02005-9

Genspect (2021). Retrieved December 20, 2021, from https://genspect.org/groups/

Getahun, D., Nash, R., Flanders, W. D., Baird, T. C., Becerra-Culqui, T. A., Cromwell, L., Hunkeler, E., Lash, T. L., Millman, A., Quinn, V. P., Robinson, B., Roblin, D., Silverberg, M. J., Safer, J., Slovis, J., Tangpricha, V., & Goodman, M. (2018). Cross-sex hormones and acute cardiovascular events in transgender persons: A cohort study. *Annals of Internal Medicine*, *169*(4), 205. doi:10.7326/M17-2785

Gooren, L., & Delemarre-van de Waal, H. (1996). The feasibility of endocrine interventions in juvenile transsexuals. *Journal of Psychology & Human Sexuality*, *8*(4), 69–74. doi:10.1300/J056v08n04_05

Green, A. E., DeChants, J. P., Price, M. N., & Davis, C. K. (2021). Association of gender-affirming hormone therapy with depression, thoughts of suicide, and attempted suicide among transgender and nonbinary youth. *Journal of Adolescent Health*, S1054139X21005681. doi:10.1016/j.jadohealth.2021.10.036

Grootens-Wiegers, P., Hein, I. M., van den Broek, J. M., & de Vries, M. C. (2017). Medical decision-making in children and adolescents: Developmental and neuroscientific aspects. *BMC Pediatrics*, *17*(1), 120. doi:10.1186/s12887-017-0869-x

Hakeem, A., Črnčec, R., Asghari-Fard, M., Harte, F., & Eapen, V. (2016). Development and validation of a measure for assessing gender dysphoria in adults: The Gender Preoccupation and Stability Questionnaire. *International Journal of Transgenderism*, *17*(3–4), 131–140. doi:10.1080/15532739.2016.1217812

Hall, R., Mitchell, L., & Sachdeva, J. (2021). Access to care and frequency of detransition among a cohort discharged by a UK national adult gender identity clinic: Retrospective case-note review. *BJPsych Open*, *7*(6), e184. doi:10.1192/bjo.2021.1022

Hembree, W. C., Cohen-Kettenis, P. T., Gooren, L., Hannema, S. E., Meyer, W. J., Murad, M. H., Rosenthal, S. M., Safer, J. D., Tangpricha, V., & T'Sjoen, G. G. (2017). Endocrine treatment of gender-dysphoric/gender-incongruent persons: An endocrine society clinical practice guideline. J Clin Endocrinol Metab, *102*(11), 3869–3903. doi:10.1210/jc.2017-01658

HHS. (2021). What does "suicide contagion" mean, and what can be done to prevent it? Retrieved December 28, 2021, from https://www.hhs.gov/answers/mental-health-and-substance-abuse/what-does-suicide-contagion-mean/index.html

HRC. (n.d.). Clinical care for gender-expansive children & adolescents. Retrieved January 4, 2022, from https://www.hrc.org/resources/interactive-map-clinical-care-programs-for-gender-nonconforming-childr

Hutchinson, A., Midgen, M., & Spiliadis, A. (2020). In support of research into rapid-onset gender dysphoria. *Archives of Sexual Behavior*, *49*(1), 79–80. doi:10.1007/s10508-019-01517-9

Iliadis, S. I., Axfors, C., Friberg, A., Arinell, H., Beckman, U., Fazekas, A., Frisen, L., Sandström, L., Thelin, N., Wahlberg, J., Södersten, M., & Papadopoulos, F. C. (2020). Psychometric properties and concurrent validity

of the Transgender Congruence Scale (TCS) in the Swedish setting. *Scientific Reports*, *10*(1), 18701. doi:10.1038/s41598-020-73663-3

James, S. E., Herman, J. L., Rankin, S., Keisling, M., Mottet, L., & Anafi, M. (2016). *The report of the 2015 U.S. Transgender Survey*. Washington, DC: National Center for Transgender Equality.

Johns, M. M., Lowry, R., Andrzejewski, J., Barrios, L. C., Demissie, Z., McManus, T., Rasberry, C. N., Robin, L., & Underwood, J. M. (2019). Transgender identity and experiences of violence victimization, substance use, suicide risk, and sexual risk behaviors among high school students – 19 states and large urban school districts, 2017. *MMWR. Morbidity and Mortality Weekly Report*, *68*(3), 67–71. doi:10.15585/mmwr.mm6803a3

Kaltiala, R., Heino, E., Työläjärvi, M., & Suomalainen, L. (2020). Adolescent development and psychosocial functioning after starting cross-sex hormones for gender dysphoria. *Nordic Journal of Psychiatry*, *74*(3), 213–219. doi:10.1080/08039488.2019.1691260

Kaltiala-Heino, R., Bergman, H., Työläjärvi, M., & Frisen, L. (2018). Gender dysphoria in adolescence: Current perspectives. *Adolescent Health, Medicine and Therapeutics*, Volume *9*, 31–41. doi:10.2147/AHMT.S135432

Kaltiala-Heino, R., Sumia, M., Työläjärvi, M., & Lindberg, N. (2015). Two years of gender identity service for minors: Overrepresentation of natal girls with severe problems in adolescent development. *Child and Adolescent Psychiatry and Mental Health*, *9*(1), 9. doi:10.1186/s13034-015-0042-y

Katz, A. L., Macauley, R. C., Mercurio, M. R., Moon, M. R., Okun, A. L., Opel, D. J., & Statter, M. B. (2016). Informed consent in decision-making in pediatric practice. Committee on Bioethics. *Pediatrics*, *138*(2), e20161484. doi:10.1542/peds.2016-1484

Kendler K. S. (2019). From many to one to many-the search for causes of psychiatric illness. *JAMA psychiatry*, *76*(10), 1085–1091. doi:10.1001/jamapsychiatry.2019.1200

Kidd, K. M., Sequeira, G. M., Douglas, C., Paglisotti, T., Inwards-Breland, D. J., Miller, E., & Coulter, R. W. S. (2021). Prevalence of gender-diverse youth in an urban school district. *Pediatrics*, *147*(6), e2020049823. doi:10.1542/peds.2020-049823

Korte, A., Goecker, D., Krude, H., Lehmkuhl, U., Grüters-Kieslich, A., & Beier, K. M. (2008). Gender identity disorders in childhood and adolescence: Currently debated concepts and treatment strategies. *Deutsches Ärzteblatt International*, *105*(48), 834–841. doi:10.3238/arztebl.2008.0834

Kozlowska, K., Chudleigh, C., McClure, G., Maguire, A. M., & Ambler, G. R. (2021). Attachment patterns in children and adolescents with gender dysphoria. *Frontiers in Psychology*, *11*. doi:10.3389/fpsyg.2020.582688

Kozlowska, K., McClure, G., Chudleigh, C., Maguire, A. M., Gessler, D., Scher, S., & Ambler, G. R. (2021). Australian children and adolescents with gender dysphoria: Clinical presentations and challenges experienced by a multidisciplinary team and gender service. *Human Systems*, 26344041211010776. doi:10.1177/26344041211010777

Laidlaw, M. K., Van Meter, Q. L., Hruz, P. W., Van Mol, A., & Malone, W. J. (2019). Letter to the Editor: "Endocrine treatment of gender-dysphoric/gender-incongruent persons: An Endocrine Society Clinical Practice Guideline"." *The Journal of Clinical Endocrinology & Metabolism*, *104*(3), 686–687. doi:10.1210/jc.2018-01925

Levine, S. B. (2021). Reflections on the clinician's role with individuals who self-identify as transgender. *Archives of Sexual Behavior*, *50*, 3527–3536. doi:10.1007/s10508-021-02142-1

Levine, S.B. (2019). Informed Consent for Transgendered Patients, *Journal of Sex and Marital Therapy*, *45*(3):218–229. doi:10.1080/0092623X.2018.1518885

Littman, L. (2018). Parent reports of adolescents and young adults perceived to show signs of a rapid onset of gender dysphoria. *PLoS ONE 13*(8): e0202330. doi:10.1371/journal.pone.0202330

Littman, L. (2020). The use of methodologies in Littman (2018) is consistent with the use of methodologies in other studies contributing to the field of gender Dysphoria research: Response to Restar (2019). *Archives of Sexual Behavior*, *49*(1), 67–77. doi:10.1007/s10508-020-01631-z

Littman, L. (2021). Individuals treated for gender dysphoria with medical and/or surgical transition who subsequently detransitioned: A survey of 100 detransitioners. *Archives of Sexual Behavior, 50*, 3353–3369. doi:10.1007/s10508-021-02163-w

Lynch, H.F., Joffe, S., Feldman, E. (2018). Informed consent and the role of the treating physician. NEJM *378*:25, 435–438.

Malone, W., D'Angelo, R., Beck, S., Mason, J., & Evans, M. (2021). Puberty blockers for gender dysphoria: The science is far from settled. *The Lancet Child & Adolescent Health*, *5*(9), e33–e34. doi:10.1016/S2352-4642(21)00235-2

Malone, W. J., Hruz, P. W., Mason, J. W., & Beck, S. (2021). Letter to the editor from william j. Malone et al: "Proper care of transgender and gender-diverse persons in the setting of proposed discrimination: a policy perspective"." J Clin Endocrinol Metab, *106*(8), e3287–e3288. doi:10.1210/clinem/dgab205

Manrique, O. J., Adabi, K., Martinez-Jorge, J., Ciudad, P., Nicoli, F., & Kiranantawat, K. (2018). Complications and patient-reported outcomes in male-to-female vaginoplasty–where we are today: A systematic review and meta-analysis. *Annals of Plastic Surgery*, *80*(6), 684–691. doi:10.1097/SAP.0000000000001393

Mars, B., Heron, J., Klonsky, E. D., Moran, P., O'Connor, R. C., Tilling, K., Wilkinson, P., & Gunnell, D. (2019). Predictors of future suicide attempt among adolescents with suicidal thoughts or non-suicidal self-harm: A population-based birth cohort study. *The Lancet Psychiatry*, *6*(4), 327–337. doi:10.1016/S2215-0366(19)30030-6

Mathes, T., & Pieper, D. (2017). Clarifying the distinction between case series and cohort studies in systematic reviews of comparative studies: Potential impact on body of evidence and workload. *BMC Medical Research Methodology*, *17*(1), 107. doi:10.1186/s12874-017-0391-8

McGuire, J. K., Berg, D., Catalpa, J. M., Morrow, Q. J., Fish, J. N., Nic Rider, G., Steensma, T., Cohen-Kettenis, P. T., & Spencer, K. (2020). Utrecht Gender Dysphoria Scale - gender spectrum (UGDS-GS): Construct validity among transgender, nonbinary, and LGBQ samples. *International Journal of Transgender Health*, *21*(2), 194–208. doi:10.1080/26895269.2020.1723460

Morandini, J. S., Kelly, A., de Graaf, N. M., Carmichael, P., & Dar-Nimrod, I. (2021). Shifts in demographics and mental health co-morbidities among gender dysphoric youth referred to a specialist gender dysphoria service. *Clinical Child Psychology and Psychiatry*, 135910452110468. doi:10.1177/13591045211046813

National Institute for Health and Care Excellence. (2020a). Evidence review: Gonadotrophin releasing hormone analogues for children and adolescents with gender dysphoria. https://arms.nice.org.uk/resources/hub/1070905/attachment

National Institute for Health and Care Excellence. (2020b). Evidence review: Gender-affirming hormones for children and adolescents with gender dysphoria. https://arms.nice.org.uk/resources/hub/1070871/attachment

Nota, N. M., Wiepjes, C. M., de Blok, C. J. M., Gooren, L. J. G., Kreukels, B. P. C., & den Heijer, M. (2019). Occurrence of acute cardiovascular events in transgender individuals receiving hormone therapy: Results from a Large Cohort Study. *Circulation*, *139*(11), 1461–1462. doi:10.1161/CIRCULATIONAHA.118.038584

Olson, K. R., Durwood, L., DeMeules, M., & McLaughlin, K. A. (2016). Mental health of transgender children who are supported in their identities. *Pediatrics*, *137*(3), 1–16. iii.

Ouliaris, C. (2021). Consent for treatment of gender dysphoria in minors: evolving clinical and legal frameworks. *The Medical journal of Australia*, Advance online publication. doi:10.5694/mja2.51357

Paine, E. A. (2021). "Fat broken arm syndrome": Negotiating risk, stigma, and weight bias in LGBTQ healthcare. Soc Sci Med, *270*, 113609. doi:10.1016/j.socscimed.2020.113609

Planned Parenthood League of Massachusetts. (n.d.) Gender affirming hormone therapy. Retrieved December 26, 2021, from https://www.plannedparenthood.org/planned-parenthood-massachusetts/campaigns/gender-affirming-hormone-therapy

Rafferty, J., Committee on Psychosocial Aspects of Child and Family Health, Committee on Adolescence, & Section on Lesbian, Gay, Bisexual, and Transgender Health and Wellness. (2018). Ensuring comprehensive care and support for transgender and gender-diverse children and adolescents. *Pediatrics*, *142*(4), e20182162. doi:10.1542/peds.2018-2162

Right to medical or dental treatment without parental consent, Oregon ORS Volume 3, Title 11, 109.640 (2010). https://oregon.public.law/statutes/ors_109.640

Ristori, J., & Steensma, T. D. (2016). Gender dysphoria in childhood. *International Review of Psychiatry*, *28*(1), 13–20. doi:10.3109/09540261.2015.1115754

Rood, B. A., Reisner, S. L., Surace, F. I., Puckett, J. A., Maroney, M. R., & Pantalone, D. W. (2016). Expecting rejection: Understanding the minority stress experiences of transgender and gender-nonconforming individuals. *Transgender Health*, *1*(1), 151–164. doi:10.1089/trgh.2016.0012

Ross, M. W., & Need, J. A. (1989). Effects of adequacy of gender reassignment surgery on psychological adjustment: A follow-up of fourteen male-to-female patients. *Archives of Sexual Behavior*, *18*(2), 145–153. doi:10.1007/BF01543120

Schulz, S. L. (2018). The informed consent model of transgender care: An alternative to the diagnosis of gender dysphoria. *Journal of Humanistic Psychology*, *58*(1), 72–92. doi:10.1177/0022167817745217

Simon, G. E., & VonKorff, M. (1998). Suicide mortality among patients treated for depression in an insured population. *American Journal of Epidemiology*, *147*, 155–160. doi:10.1093/oxfordjournals.aje.a009428

Singh, D., Bradley, S. J., & Zucker, K. J. (2021). A follow-up study of boys with gender identity disorder. *Frontiers in Psychiatry*, *12*. doi:10.3389/fpsyt.2021.632784

Smith, Y. L. S., Van Goozen, S. H. M., & Cohen-Kettenis, P. T. (2001). Adolescents with gender identity disorder who were accepted or rejected for sex reassignment surgery: A prospective Follow-up Study. J Am Acad Child Adolesc Psychiatry, *40*(4), 472–481. doi:10.1097/00004583-200104000-00017

Smith, A. R., Zuromski, K. L., & Dodd, D. R. (2018). Eating disorders and suicidality: What we know, what we don't know, and suggestions for future research. *Current Opinion in Psychology*, *22*, 63–67. doi:10.1016/j.copsyc.2017.08.023

Spiliadis, A. (2019). Towards a gender exploratory model: Slowing things down, opening things up and exploring identity development. Metalogos Systemic Therapy Journal, *35*, 1–9. https://www.ohchr.org/Documents/Issues/SexualOrientation/IESOGI/Other/Rebekah_Murphy_TowardsaGenderExploratoryModelslowingthingsdownopeningthingsupandexploringidentitydevelopment.pdf

Steensma, T. D., McGuire, J. K., Kreukels, B. P. C., Beekman, A. J., & Cohen-Kettenis, P. T. (2013). Factors associated with desistence and persistence of childhood gender dysphoria: A quantitative follow-up study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(6), 582–590. doi:10.1016/j.jaac.2013.03.016

Strang, J. F., Janssen, l, Tishelman, A., Leibowitz, S. F., Kenworthy, L., McGuire, J. K., Edwards-Leeper, L., Mazefsky, C. A., Rofey, D., Bascom, J., Caplan, R., Gomez-Lobo, V., Berg, D., Zaks, Z., Wallace, G. L., Wimms, H., Pine-Twaddell, E., Shumer, D., Register-Brown, K., … Anthony, L. G. (2018). Revisiting the link: Evidence of

the rates of autism in studies of gender diverse individuals. *Journal of the American Academy of Child and Adolescent Psychiatry*, *57*(11), 885–887. doi:10.1016/j.jaac.2018.04.023

Tavistock and Portman NHS Foundation Trust. (2020). Gender identity development service referrals in 2019–20 same as 2018–19. https://tavistockandportman.nhs.uk/about-us/news/stories/gender-identity-development-service-referrals-2019-20-same-2018-19/

Temple Newhook, J., Pyne, J., Winters, K., Feder, S., Holmes, C., Tosh, J., Sinnott, M.-L., Jamieson, A., & Pickett, S. (2018). A critical commentary on follow-up studies and "desistance" theories about transgender and gender-nonconforming children. *International Journal of Transgenderism*, *19*(2), 212–224. doi:10.1080/15532739.2018.1456390

The Trevor Project. (2021). National Survey on LGBTQ Youth Mental Health 2021. Retrieved January 3, 2022, from https://www.thetrevorproject.org/survey-2021/?section=SuicideMentalHealth

Toomey, R. B., Syvertsen, A. K., & Shramko, M. (2018). Transgender Adolescent Suicide Behavior. *Pediatrics*, *142*(4). doi:10.1542/peds.2017-4218

Turban, J. L. (2018). Potentially reversible social deficits among transgender youth. *Journal of Autism and Developmental Disorders*, *48*(12), 4007–4009. doi:10.1007/s10803-018-3603-0

Turban, J. L., King, D., Carswell, J. M., & Keuroghlian, A. S. (2020). Pubertal suppression for transgender youth and risk of suicidal ideation. *Pediatrics*, *145*(2), e20191725. doi:10.1542/peds.2019-1725

Turban, J. L., & van Schalkwyk, G. I. (2018). "Gender Dysphoria" and autism spectrum disorder: is the link real? *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*(1), 8–9.e2. doi:10.1016/j.jaac.2017.08.017

van der Miesen, A. I. R., Cohen-Kettenis, P. T., & de Vries, A. L. C. (2018). Is there a link between gender dysphoria and autism spectrum disorder? *Journal of the American Academy of Child & Adolescent Psychiatry*, *57*(11), 884–885. doi:10.1016/j.jaac.2018.04.022

Vandenbussche, E. (2021). Detransition-related needs and support: A cross-sectional online survey. *Journal of Homosexuality*, *20,* 1–19. doi:10.1080/00918369.2021.1919479

Voorzij. (2021). More research is urgently needed into transgender care for young people: "Where does the large increase of children come from?" Retrieved December 20, 2021, from https://www.voorzij.nl/more-research-is-urgently-needed-into-transgender-care-for-young-people-where-does-the-large-increase-of-children-come-from/.

Vrouenraets, L., de Vries, A., de Vries, M. C., van der Miesen, A., & Hein, I. M. (2021). Assessing medical decision-making competence in transgender youth. *Pediatrics*, *148*, e2020049643. Advance online publication. doi:10.1542/peds.2020-049643

Vrouenraets, L., Hartman, L. A., Hein, I. M., de Vries, A., de Vries, M. C., & Molewijk, B. (2020). Dealing with moral challenges in treatment of transgender children and adolescents: Evaluating the role of moral case deliberation. *Archives of sexual behavior*, *49*(7), 2619–2634. doi:10.1007/s10508-020-01762-3

Wiepjes, C. M., den Heijer, M., Bremmer, M. A., Nota, N. M., Blok, C. J. M., Coumou, B. J. G., & Steensma, T. D. (2020). Trends in suicide death risk in transgender people: Results from the Amsterdam Cohort of Gender Dysphoria Study (1972–2017)). *Acta Psychiatrica Scandinavica*, *141*(6), 486–491. doi:10.1111/acps.13164

Wiepjes, C. M., Nota, N. M., de Blok, C. J. M., Klaver, M., de Vries, A. L. C., Wensing-Kruger, S. A., de Jongh, R. T., Bouman, M.-B., Steensma, T. D., Cohen-Kettenis, P., Gooren, L. J. G., Kreukels, B. P. C., & den Heijer, M. (2018). The Amsterdam Cohort of Gender Dysphoria Study (1972–2015): Trends in Prevalence, Treatment, and Regrets. *The Journal of Sexual Medicine*, *15*(4), 582–590. doi:10.1016/j.jsxm.2018.01.016

Wilson, S. C., Morrison, S. D., Anzai, L., Massie, J. P., Poudrier, G., Motosko, C. C., & Hazen, A. (2018). Masculinizing top surgery: A systematic review of techniques and outcomes. *Annals of Plastic Surgery*, *80*(6), 679–683. doi:10.1097/SAP.0000000000001354

World Health Organization. (2019). International statistical classification of diseases and related health problems (11th ed.). https://icd.who.int/

Zucker, K. J. (2017). Epidemiology of gender dysphoria and transgender identity. *Sexual Health*, *14*(5), 404. doi:10.1071/SH17067

Zucker, K. J. (2018). The myth of persistence: Response to "A critical commentary on follow-up studies and 'desistance' theories about transgender and gender non-conforming children" by Temple Newhook et al. (2018). *International Journal of Transgenderism*, *19*(2), 231–245. doi:10.1080/15532739.2018.1468293

Zucker, K. J. (2019). Adolescents with gender dysphoria: Reflections on some contemporary clinical and research issues. *Archives of Sexual Behavior*, *48*(7), 1983–1992. doi:10.1007/s10508-019-01518-8

Zucker, K. J. (2020). Debate: Different strokes for different folks. *Child and Adolescent Mental Health*, *25*(1), 36–37. doi:10.1111/camh.12330

Reproductive Endocrinology: *Case Report*

Pregnancy Outcomes After Fertility Preservation in Transgender Men

Susan Maxwell, MD,
Nicole Noyes, MD,
David Keefe, MD,
Alan S. Berkeley, MD,
and Kara N. Goldman, MD

BACKGROUND: Transgender individuals, individuals whose gender identity does not align with their sex assigned at birth, undergoing gender-affirming hormonal or surgical therapies may experience loss of fertility. Assisted reproductive technologies have expanded family-building options for transgender men who were assigned female at birth.

CASES: Three transgender men underwent oocyte cryopreservation before gender-affirming hormonal therapy. One patient underwent fertility preservation as an adolescent. Two adult patients had children using their cryopreserved oocytes, with the pregnancies carried by their sexually intimate partners.

CONCLUSION: Transgender men with cryopreserved gametes can build families in a way that affirms their gender identity. Obstetrician–gynecologists should be familiar with the fertility needs of transgender patients so appropriate discussions and referrals can be made.

(Obstet Gynecol 2017;129:1031–4)
DOI: 10.1097/AOG.0000000000002036

Teaching Points

1. Transgender individuals considering gender-affirming therapies should be counseled about fertility preservation.
2. Many family-building options using assisted reproductive technologies are available for transgender men.

Gender nonconformity occurs when an individual's gender identity differs from societal expectations. Transgender is an umbrella term that refers to individuals with a gender identity that does not align with their sex assigned at birth; therefore, transgender men are men who were assigned female at birth. Cisgender individuals' gender identity aligns with their sex assigned at birth. Gender dysphoria is psychological distress caused by a discrepancy between a person's gender identity and their sex assigned at birth.[1] Treatment for gender dysphoria may include psychotherapy and gender-affirming hormone and surgical therapies, which may result in loss of fertility. The World Professional Association for Transgender Health publishes guidelines for health care professionals caring for transgender individuals. They urge physicians to discuss the effect of gender-affirming treatments on fertility.[1] Physicians should be familiar with the fertility preservation options available for transgender patients.

This case series presents three transgender men who underwent oocyte cryopreservation before initiation of gender-affirming hormone therapy. Two patients returned to use their gametes and had pregnancies carried by their sexually intimate partners. Patients' preferred pronouns (ie, he and his) are used. All patients provided written permission to include their deidentified medical histories. Successful oocyte cryopreservation in an adolescent transgender boy has been documented previously; however, our case series uniquely describes the family-building options available to transgender individuals after fertility preservation.[2]

CASE 1

The patient is a 17-year-old adolescent transgender boy, gravida 0, referred by his endocrinologist for fertility preservation before initiating androgen therapy. He was accompanied by his mother, who supported his decision to transition before attending college. After counseling, he

From the Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, New York University School of Medicine, New York, New York.

The authors thank the patients of NYUFC for permitting us to share their fertility preservation stories and Tiffany Cook for her expertise and review of our manuscript.

Each author has indicated that he or she has met the journal's requirements for authorship.

Corresponding author: Susan Maxwell, MD, Division of Reproductive Endocrinology and Infertility, Department of Obstetrics and Gynecology, New York University School of Medicine, 660 First Avenue, 5th Floor, New York, NY 10016; email: Susan.mcclelland@nyumc.org.

Financial Disclosure
The authors did not report any potential conflicts of interest.

ISSN: 0029-7844/17

elected to undergo oocyte cryopreservation. Controlled ovarian hyperstimulation was performed in a routine fashion using a gonadotropin-releasing hormone antagonist protocol. A total of 21 oocytes were retrieved and 17 mature oocytes were cryopreserved with vitrification. His oocytes remain in storage.

CASE 2

The patient is a 30-year-old single transgender man, gravida 0, who presented for fertility preservation with oocyte cryopreservation before initiating androgen therapy. Controlled ovarian hyperstimulation was performed using a low-dose leuprolide acetate protocol. Forty-five mature oocytes were obtained and cryopreserved.

After 8 years on androgen therapy, he returned with his partner, a 34-year-old cisgender woman, gravida 0. He desired embryo creation using donor sperm with transfer of embryos into his cisgender partner. On pan-ethnic carrier screening, he was found to carry a fragile X premutation. His partner underwent physical examination, saline infusion ultrasonography, and U.S. Food and Drug Administration testing, which were normal. The couple requested preimplantation genetic screening and diagnosis for aneuploidy and fragile X.

Twenty of the 45 oocytes were thawed, and 18 (90%) survived, with 17 (94%) fertilized after intracytoplasmic sperm injection. The embryos were cultured for 5–6 days, and 12 blastocysts underwent trophectoderm biopsy followed by vitrification. Comprehensive chromosomal screening using array comparative genomic hybridization resulted in six euploid embryos. Five of those were fragile X premutation carriers and one was unaffected. The endometrium was prepared using sequentially increasing doses of oral estradiol until the diameter reached 8.4 mm. Fifty milligrams of daily intramuscular progesterone in oil was started on cycle day 14. The unaffected euploid embryo was thawed and transferred on the sixth day of progesterone therapy.

On cycle day 28, her serum quantitative human chorionic gonadotropin level was 180 milli-international units/mL. First-trimester ultrasonography revealed monozygotic–diamniotic twins. The twins were delivered by cesarean at 34 weeks of gestation as a result of preterm prelabor rupture of membranes and weighed 4 pounds 13 ounces and 4 pounds 9 ounces. The infants were doing well at 4-month follow-up.

CASE 3

The patient is a 32-year-old single transgender man, gravida 0, who presented for fertility preservation with oocyte cryopreservation. He had undergone a bilateral mastectomy and planned on initiating gender-affirming hormone therapy. Controlled ovarian hyperstimulation was performed using a low-dose leuprolide acetate protocol. Thirteen mature and six immature oocytes were retrieved and cryopreserved.

Five years later he desired creation of embryos with donor sperm and transfer of the embryos into his partner, a 45-year-old healthy cisgender woman, gravida 1 para 0010. He had been receiving androgen therapy. His partner underwent physical examination, saline infusion ultrasonography, and U.S. Food and Drug Administration testing. Endometrial preparation was performed as described previously and the diameter reached 12.5 mm. Vaginal progesterone twice daily was added on the day of oocyte thaw.

Nineteen cryopreserved oocytes were thawed with 95% survival. Fourteen were mature and 11 (79%) fertilized after intracytoplasmic sperm injection. Embryos were cultured for 5–6 days, one grade 4 Bb day 5 embryo was transferred, and the remaining five underwent vitrification. A biochemical pregnancy resulted. Two months later a frozen embryo transfer cycle was initiated with endometrial preparation. Two day 5 grade 4 Bb embryos were transferred on the sixth day of vaginal progesterone. Cycle day 28 and day 35 human chorionic gonadotropin levels were 263 and 4,384 milli-international units/mL. First-trimester ultrasonography revealed dichorionic–diamniotic twins. The twins were delivered by cesarean for preeclampsia at 35 5/7 weeks of gestation. One neonate weighed 4 pounds 9 ounces, and the second weighed 4 pounds 15 ounces. The infants were doing well at 11-month follow-up.

DISCUSSION

Assisted reproductive technologies have greatly expanded family-building options for transgender individuals. These cases demonstrate how transgender men with cryopreserved gametes can build families in a way that affirms their gender identity. These patients were young at the time of oocyte cryopreservation and were referred for fertility preservation before undergoing gender-affirming hormonal therapy, which contributed to their successful outcomes. Obstetrician–gynecologists should be familiar with the fertility preservation options for transgender individuals and the benefits of early fertility preservation before gender-affirming therapies.

Many transgender individuals desire a family. In a survey of 50 transgender individuals in Belgium, 54% were interested in family-building. Of 11 participants with children, eight had partners who conceived with donor sperm and three conceived themselves before undergoing gender-affirming therapies. None had used assisted reproductive technology, but 37.5% would have considered fertility preservation if available.[3]

Gender-affirming hormone therapy significantly affects reproduction. Among transgender women, estrogen therapy suppresses gonadotropin levels resulting in impaired testosterone production. Reduced testosterone levels decrease sperm count and motility. Cessation of estrogen therapy may reverse these effects, but this is not well studied. Among transgender men, testosterone therapy induces amenorrhea by suppressing ovulation. The effects of testosterone on ovarian function may also be reversible after discontinuing testosterone.

To avoid discontinuing gender-affirming hormones and the potential for compromised fertility, transgender individuals who want to maintain fertility may undergo cryopreservation of their gametes. Transgender women may cryopreserve sperm.[4] Transgender men have the option of oocyte cryopreservation or embryo banking (cryopreservation of embryos for future use) (Table 1).[2] The creation of embryos for fertility preservation using partner or donor sperm may result in improved pregnancy rates at some centers compared with the use of cryopreserved oocytes. This may have implications for the early use of donor sperm for some individuals.[5]

Gender-affirming surgical procedures involving reproductive organs cause permanent loss of fertility. Ovarian tissue cryopreservation is available for individuals undergoing oophorectomy; however, it is still considered experimental. Use of autologous transplanted ovarian tissue requires a patient to discontinue androgen therapy, and in vitro fertilization is often necessary. The number of ooctyes retrieved from stimulated ovarian grafts is variable and dependent on graft function, graft size, and ovarian reserve before tissue harvesting.[6] Among transgender oophorectomized patients, frozen–thawed ovarian tissue must be grafted to a nonovary surface, which is associated with diminished pregnancy success.

Transgender men who want the option to carry a pregnancy should consider oophorectomy without hysterectomy. Ovaries are not required to maintain a pregnancy because exogenous hormones are administered. Transgender men can carry pregnancies after stopping testosterone therapy; however, discontinuing gender-affirming hormones may cause distress.[7] Alternatively, pregnancies may be carried by a partner or gestational carrier. Transgender men undergoing oophorectomy should be counseled about the requirement for long-term access to androgen or hormone therapy to prevent bone loss associated with menopause.

Gamete cryopreservation should be considered for adolescent transgender or gender-nonconforming youth initiating gender-affirming hormones. Early initiation of gender-affirming hormones prevents irreversible physical changes and reduces gender dysphoria. Gonadotropin-releasing hormone agonists are used once adolescents reach Tanner stages 2–3 of pubertal development to suppress secondary sex characteristics. This reversible therapy allows adolescents to explore their gender identity before starting gender-affirming hormones. The Endocrine Society recommends postponing gender-affirming hormones until age 16 years and surgical procedures until age 18 years, but this remains controversial.[4]

Fertility preservation among adolescent transgender boys is predominantly restricted to oocyte cryopreservation. Adolescents respond well and tolerate controlled ovarian hyperstimulation during oocyte cryopreservation. Follicular monitoring with abdominal ultrasonograms in both adolescents and adults decreases discomfort and anxiety. Given its experimental nature, ovarian tissue cryopreservation is reserved for adolescents with a concomitant medical condition such as those undergoing gonadotoxic cancer treatments.

Table 1. Fertility Preservation Options for Adolescent Transgender Boys and Adult Transgender Men

Age Group (y)	Method	Advantages	Disadvantages
Younger than 18	OC	Well-tolerated Minimally invasive outpatient procedure	Must be perimenarchal or postmenarchal
	OTC	Available before menarche	Experimental Not recommended unless concomitant medical condition (ie, undergoing chemotherapy for cancer) Requires invasive surgical procedure
18 or older	OC	More flexibility for the future use of gametes	Difficult to estimate number of oocytes needed for a live birth Embryo formation rates may be lower at some centers
	OTC	No need to stop androgen therapy before surgery Performed at the time of planned oophorectomy	Experimental Suboptimal graft function if tissue transplantation onto peritoneal surface, requires IVF
	EB	More accurate estimate of chance of live birth Can perform CCS before embryo cryopreservation	Requires sperm source (donor or partner) Legal implications if partners separate after creating embryos

OC, oocyte cryopreservation; OTC, ovarian tissue cryopreservation; IVF, in vitro fertilization; EB, embryo banking; CCS, comprehensive chromosomal screening.

Transgender individuals face many barriers to health care and have difficulty finding physicians familiar with their medical needs.[8] Furthermore, the cost of gender-affirming therapies and fertility preservation is rarely covered by insurance. Obstetrician–gynecologists play an important role in the reproductive health of transgender patients. They must be familiar with the types of gender-affirming therapies and fertility needs of transgender individuals so appropriate counseling and referrals are made.[8]

REFERENCES

1. Coleman E, Bockting W, Botzer M, Cohen-Kettenis P, DeCuypere G, Feldman J, et al. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. Int J Transgend 2012;13:165–232.
2. Wallace SA, Blough KL, Kondapalli LA. Fertility preservation in the transgender patient: expanding oncofertility care beyond cancer. Gynecol Endocrinol 2014;30:868–71.
3. Wierckx K, Van Caenegem E, Pennings G, Elaut E, Dedecker D, Van de Peer F, et al. Reproductive wish in transsexual men. Hum Reprod 2012;27:483–7.
4. Hembree WC, Cohen-Kettenis P, Delemarre-van de Waal HA, Gooren LJ, Meyer WJ 3rd, Spack NP, et al. Endocrine treatment of transsexual persons: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab 2009;94:3132–54.
5. Practice Committees of American Society for Reproductive Medicine; Society for Assisted Reproductive Technology. Mature oocyte cryopreservation: a guideline. Fertil Steril 2013;99:37–43.
6. Meirow D, Ra'anani H, Shapira M, Brenghausen M, Derech Chaim S, Aviel-Ronen S, et al. Transplantations of frozen thawed ovarian tissue demonstrate high reproductive performance and the need to revise restrictive criteria. Fertil Steril 2016;106:467–74.
7. Light AD, Obedin-Maliver J, Sevelius JM, Kerns JL. Transgender men who experienced pregnancy after female-to-male gender transitioning. Obstet Gynecol 2014;124:1120–7.
8. Care for transgender adolescents. Committee Opinion No. 685. American College of Obstetricians and Gynecologists. Obstet Gynecol 2017;129:e11–6.

Endocrine Treatment of Gender-Dysphoric/ Gender-Incongruent Persons: An Endocrine Society* Clinical Practice Guideline

Wylie C. Hembree,[1] Peggy T. Cohen-Kettenis,[2] Louis Gooren,[3] Sabine E. Hannema,[4] Walter J. Meyer,[5] M. Hassan Murad,[6] Stephen M. Rosenthal,[7] Joshua D. Safer,[8] Vin Tangpricha,[9] and Guy G. T'Sjoen[10]

[1]New York Presbyterian Hospital, Columbia University Medical Center, New York, New York 10032 (Retired); [2]VU University Medical Center, 1007 MB Amsterdam, Netherlands (Retired); [3]VU University Medical Center, 1007 MB Amsterdam, Netherlands (Retired); [4]Leiden University Medical Center, 2300 RC Leiden, Netherlands; [5]University of Texas Medical Branch, Galveston, Texas 77555; [6]Mayo Clinic Evidence-Based Practice Center, Rochester, Minnesota 55905; [7]University of California San Francisco, Benioff Children's Hospital, San Francisco, California 94143; [8]Boston University School of Medicine, Boston, Massachusetts 02118; [9]Emory University School of Medicine and the Atlanta VA Medical Center, Atlanta, Georgia 30322; and [10]Ghent University Hospital, 9000 Ghent, Belgium

*Cosponsoring Associations: American Association of Clinical Endocrinologists, American Society of Andrology, European Society for Pediatric Endocrinology, European Society of Endocrinology, Pediatric Endocrine Society, and World Professional Association for Transgender Health.

Objective: To update the "Endocrine Treatment of Transsexual Persons: An Endocrine Society Clinical Practice Guideline," published by the Endocrine Society in 2009.

Participants: The participants include an Endocrine Society–appointed task force of nine experts, a methodologist, and a medical writer.

Evidence: This evidence-based guideline was developed using the Grading of Recommendations, Assessment, Development, and Evaluation approach to describe the strength of recommendations and the quality of evidence. The task force commissioned two systematic reviews and used the best available evidence from other published systematic reviews and individual studies.

Consensus Process: Group meetings, conference calls, and e-mail communications enabled consensus. Endocrine Society committees, members and cosponsoring organizations reviewed and commented on preliminary drafts of the guidelines.

Conclusion: Gender affirmation is multidisciplinary treatment in which endocrinologists play an important role. Gender-dysphoric/gender-incongruent persons seek and/or are referred to endocrinologists to develop the physical characteristics of the affirmed gender. They require a safe and effective hormone regimen that will (1) suppress endogenous sex hormone secretion determined by the person's genetic/gonadal sex and (2) maintain sex hormone levels within the normal range for the person's affirmed gender. Hormone treatment is not recommended for prepubertal gender-dysphoric/gender-incongruent persons. Those clinicians who recommend gender-affirming endocrine treatments—appropriately trained diagnosing clinicians (required), a mental health provider for adolescents (required) and mental health

ISSN Print 0021-972X ISSN Online 1945-7197
Printed in USA

Received 24 July 2017. Accepted 24 August 2017.
First Published Online 13 September 2017

Abbreviations: BMD, bone mineral density; DSD, disorder/difference of sex development; DSM, Diagnostic and Statistical Manual of Mental Disorders; GD, gender dysphoria; GnRH, gonadotropin-releasing hormone; ICD, International Statistical Classification of Diseases and Related Health Problems; MHP, mental health professional; VTE, venous thromboembolism.

professional for adults (recommended)—should be knowledgeable about the diagnostic criteria and criteria for gender-affirming treatment, have sufficient training and experience in assessing psychopathology, and be willing to participate in the ongoing care throughout the endocrine transition. We recommend treating gender-dysphoric/gender-incongruent adolescents who have entered puberty at Tanner Stage G2/B2 by suppression with gonadotropin-releasing hormone agonists. Clinicians may add gender-affirming hormones after a multidisciplinary team has confirmed the persistence of gender dysphoria/gender incongruence and sufficient mental capacity to give informed consent to this partially irreversible treatment. Most adolescents have this capacity by age 16 years old. We recognize that there may be compelling reasons to initiate sex hormone treatment prior to age 16 years, although there is minimal published experience treating prior to 13.5 to 14 years of age. For the care of peripubertal youths and older adolescents, we recommend that an expert multidisciplinary team comprised of medical professionals and mental health professionals manage this treatment. The treating physician must confirm the criteria for treatment used by the referring mental health practitioner and collaborate with them in decisions about gender-affirming surgery in older adolescents. For adult gender-dysphoric/gender-incongruent persons, the treating clinicians (collectively) should have expertise in transgender-specific diagnostic criteria, mental health, primary care, hormone treatment, and surgery, as needed by the patient. We suggest maintaining physiologic levels of gender-appropriate hormones and monitoring for known risks and complications. When high doses of sex steroids are required to suppress endogenous sex steroids and/or in advanced age, clinicians may consider surgically removing natal gonads along with reducing sex steroid treatment. Clinicians should monitor both transgender males (female to male) and transgender females (male to female) for reproductive organ cancer risk when surgical removal is incomplete. Additionally, clinicians should persistently monitor adverse effects of sex steroids. For gender-affirming surgeries in adults, the treating physician must collaborate with and confirm the criteria for treatment used by the referring physician. Clinicians should avoid harming individuals (via hormone treatment) who have conditions other than gender dysphoria/gender incongruence and who may not benefit from the physical changes associated with this treatment. (***J Clin Endocrinol Metab*** **102: 3869–3903, 2017)**

Summary of Recommendations

1.0 Evaluation of youth and adults

1.1. We advise that only trained mental health professionals (MHPs) who meet the following criteria should diagnose gender dysphoria (GD)/gender incongruence in adults: (1) competence in using the Diagnostic and Statistical Manual of Mental Disorders (DSM) and/or the International Statistical Classification of Diseases and Related Health Problems (ICD) for diagnostic purposes, (2) the ability to diagnose GD/gender incongruence and make a distinction between GD/gender incongruence and conditions that have similar features (*e.g.*, body dysmorphic disorder), (3) training in diagnosing psychiatric conditions, (4) the ability to undertake or refer for appropriate treatment, (5) the ability to psychosocially assess the person's understanding, mental health, and social conditions that can impact gender-affirming hormone therapy, and (6) a practice of regularly attending relevant professional meetings. (Ungraded Good Practice Statement)

1.2. We advise that only MHPs who meet the following criteria should diagnose GD/gender incongruence in children and adolescents: (1) training in child and adolescent developmental psychology and psychopathology, (2) competence in using the DSM and/or the ICD for diagnostic purposes, (3) the ability to make a distinction between GD/gender incongruence and conditions that have similar features (*e.g.*, body dysmorphic disorder), (4) training in diagnosing psychiatric conditions, (5) the ability to undertake or refer for appropriate treatment, (6) the ability to psychosocially assess the person's understanding and social conditions that can impact gender-affirming hormone therapy, (7) a practice of regularly attending relevant professional meetings, and (8) knowledge of the criteria for puberty blocking and gender-affirming hormone treatment in adolescents. (Ungraded Good Practice Statement)

1.3. We advise that decisions regarding the social transition of prepubertal youths with GD/gender incongruence are made with the assistance of an MHP or another experienced professional. (Ungraded Good Practice Statement)

doi: 10.1210/jc.2017-01658

1.4. We recommend against puberty blocking and gender-affirming hormone treatment in prepubertal children with GD/gender incongruence. (1 |⊕⊕◯◯)

1.5. We recommend that clinicians inform and counsel all individuals seeking gender-affirming medical treatment regarding options for fertility preservation prior to initiating puberty suppression in adolescents and prior to treating with hormonal therapy of the affirmed gender in both adolescents and adults. (1 |⊕⊕⊕◯)

2.0 Treatment of adolescents

2.1. We suggest that adolescents who meet diagnostic criteria for GD/gender incongruence, fulfill criteria for treatment, and are requesting treatment should initially undergo treatment to suppress pubertal development. (2 |⊕⊕◯◯)

2.2. We suggest that clinicians begin pubertal hormone suppression after girls and boys first exhibit physical changes of puberty. (2 |⊕⊕◯◯)

2.3. We recommend that, where indicated, GnRH analogues are used to suppress pubertal hormones. (1 |⊕⊕◯◯)

2.4. In adolescents who request sex hormone treatment (given this is a partly irreversible treatment), we recommend initiating treatment using a gradually increasing dose schedule after a multidisciplinary team of medical and MHPs has confirmed the persistence of GD/gender incongruence and sufficient mental capacity to give informed consent, which most adolescents have by age 16 years. (1 |⊕⊕◯◯).

2.5. We recognize that there may be compelling reasons to initiate sex hormone treatment prior to the age of 16 years in some adolescents with GD/gender incongruence, even though there are minimal published studies of gender-affirming hormone treatments administered before age 13.5 to 14 years. As with the care of adolescents ≥16 years of age, we recommend that an expert multidisciplinary team of medical and MHPs manage this treatment. (1 |⊕◯◯◯)

2.6. We suggest monitoring clinical pubertal development every 3 to 6 months and laboratory parameters every 6 to 12 months during sex hormone treatment. (2 |⊕⊕◯◯)

3.0 Hormonal therapy for transgender adults

3.1. We recommend that clinicians confirm the diagnostic criteria of GD/gender incongruence and the criteria for the endocrine phase of gender transition before beginning treatment. (1 |⊕⊕⊕◯)

3.2. We recommend that clinicians evaluate and address medical conditions that can be exacerbated by hormone depletion and treatment with sex hormones of the affirmed gender before beginning treatment. (1 |⊕⊕⊕◯)

3.3. We suggest that clinicians measure hormone levels during treatment to ensure that endogenous sex steroids are suppressed and administered sex steroids are maintained in the normal physiologic range for the affirmed gender. (2 |⊕⊕◯◯)

3.4. We suggest that endocrinologists provide education to transgender individuals undergoing treatment about the onset and time course of physical changes induced by sex hormone treatment. (2 |⊕◯◯◯)

4.0 Adverse outcome prevention and long-term care

4.1. We suggest regular clinical evaluation for physical changes and potential adverse changes in response to sex steroid hormones and laboratory monitoring of sex steroid hormone levels every 3 months during the first year of hormone therapy for transgender males and females and then once or twice yearly. (2 |⊕⊕◯◯)

4.2. We suggest periodically monitoring prolactin levels in transgender females treated with estrogens. (2 |⊕⊕◯◯)

4.3. We suggest that clinicians evaluate transgender persons treated with hormones for cardiovascular risk factors using fasting lipid profiles, diabetes screening, and/or other diagnostic tools. (2 |⊕⊕◯◯)

4.4. We recommend that clinicians obtain bone mineral density (BMD) measurements when risk factors for osteoporosis exist, specifically in those who stop sex hormone therapy after gonadectomy. (1 |⊕⊕◯◯)

4.5. We suggest that transgender females with no known increased risk of breast cancer follow breast-screening guidelines recommended for non-transgender females. (2 |⊕⊕◯◯)

4.6. We suggest that transgender females treated with estrogens follow individualized screening according to personal risk for prostatic disease and prostate cancer. (2 |⊕◯◯◯)

4.7. We advise that clinicians determine the medical necessity of including a total hysterectomy and oophorectomy as part of gender-affirming surgery. (Ungraded Good Practice Statement)

5.0 Surgery for sex reassignment and gender confirmation

5.1. We recommend that a patient pursue genital gender-affirming surgery only after the MHP and the clinician responsible for endocrine transition therapy both agree that surgery is medically necessary and would benefit the patient's overall health and/or well-being. (1 ⊕⊕◯◯)

5.2. We advise that clinicians approve genital gender-affirming surgery only after completion of at least 1 year of consistent and compliant hormone treatment, unless hormone therapy is not desired or medically contraindicated. (Ungraded Good Practice Statement)

5.3. We advise that the clinician responsible for endocrine treatment and the primary care provider ensure appropriate medical clearance of transgender individuals for genital gender-affirming surgery and collaborate with the surgeon regarding hormone use during and after surgery. (Ungraded Good Practice Statement)

5.4. We recommend that clinicians refer hormone-treated transgender individuals for genital surgery when: (1) the individual has had a satisfactory social role change, (2) the individual is satisfied about the hormonal effects, and (3) the individual desires definitive surgical changes. (1 ⊕◯◯◯)

5.5. We suggest that clinicians delay gender-affirming genital surgery involving gonadectomy and/or hysterectomy until the patient is at least 18 years old or legal age of majority in his or her country. (2 ⊕⊕◯◯).

5.6. We suggest that clinicians determine the timing of breast surgery for transgender males based upon the physical and mental health status of the individual. There is insufficient evidence to recommend a specific age requirement. (2 ⊕◯◯◯)

Changes Since the Previous Guideline

Both the current guideline and the one published in 2009 contain similar sections. Listed here are the sections contained in the current guideline and the corresponding number of recommendations: Introduction, Evaluation of Youth and Adults (5), Treatment of Adolescents (6), Hormonal Therapy for Transgender Adults (4), Adverse Outcomes Prevention and Long-term Care (7), and Surgery for Sex Reassignment and Gender Confirmation (6). The current introduction updates the diagnostic classification of "gender dysphoria/gender incongruence." It also reviews the development of "gender identity" and summarizes its natural development. The section on clinical evaluation of both youth and adults, defines in detail the professional qualifications required of those who diagnose and treat both adolescents and adults. We advise that decisions regarding the social transition of prepubertal youth are made with the assistance of a mental health professional or similarly experienced professional. We recommend against puberty blocking followed by gender-affirming hormone treatment of prepubertal children. Clinicians should inform pubertal children, adolescents, and adults seeking gender-confirming treatment of their options for fertility preservation. Prior to treatment, clinicians should evaluate the presence of medical conditions that may be worsened by hormone depletion and/or treatment. A multidisciplinary team, preferably composed of medical and mental health professionals, should monitor treatments. Clinicians evaluating transgender adults for endocrine treatment should confirm the diagnosis of persistent gender dysphoria/gender incongruence. Physicians should educate transgender persons regarding the time course of steroid-induced physical changes. Treatment should include periodic monitoring of hormone levels and metabolic parameters, as well as assessments of bone density and the impact upon prostate, gonads, and uterus. We also make recommendations for transgender persons who plan genital gender-affirming surgery.

Method of Development of Evidence-Based Clinical Practice Guidelines

The Clinical Guidelines Subcommittee (CGS) of the Endocrine Society deemed the diagnosis and treatment of individuals with GD/gender incongruence a priority area for revision and appointed a task force to formulate evidence-based recommendations. The task force followed the approach recommended by the Grading of Recommendations, Assessment, Development, and Evaluation group, an international group with expertise in the development and implementation of evidence-based guidelines (1). A detailed description of the grading scheme has been published elsewhere (2). The task force used the best available research evidence to develop the recommendations. The task force also used consistent language and graphical descriptions of both the strength of a recommendation and the quality of evidence. In terms of the strength of the recommendation, strong recommendations use the phrase "we recommend" and the number 1, and weak recommendations use the phrase "we suggest" and the number 2. Cross-filled circles indicate the quality of the evidence, such that ⊕◯◯◯ denotes very low–quality evidence; ⊕⊕◯◯, low quality; ⊕⊕⊕◯, moderate quality; and ⊕⊕⊕⊕, high quality. The task force has confidence that persons who receive care according to the strong recommendations will derive, on average, more benefit than harm. Weak recommendations require more careful consideration of the person's circumstances, values, and preferences to determine the best course of action. Linked to each recommendation is a description of the evidence and the

doi: 10.1210/jc.2017-01658

values that the task force considered in making the recommendation. In some instances, there are remarks in which the task force offers technical suggestions for testing conditions, dosing, and monitoring. These technical comments reflect the best available evidence applied to a typical person being treated. Often this evidence comes from the unsystematic observations of the task force and their preferences; therefore, one should consider these remarks as suggestions.

In this guideline, the task force made several statements to emphasize the importance of shared decision-making, general preventive care measures, and basic principles of the treatment of transgender persons. They labeled these "Ungraded Good Practice Statement." Direct evidence for these statements was either unavailable or not systematically appraised and considered out of the scope of this guideline. The intention of these statements is to draw attention to these principles.

The Endocrine Society maintains a rigorous conflict-of-interest review process for developing clinical practice guidelines. All task force members must declare any potential conflicts of interest by completing a conflict-of-interest form. The CGS reviews all conflicts of interest before the Society's Council approves the members to participate on the task force and periodically during the development of the guideline. All others participating in the guideline's development must also disclose any conflicts of interest in the matter under study, and most of these participants must be without any conflicts of interest. The CGS and the task force have reviewed all disclosures for this guideline and resolved or managed all identified conflicts of interest.

Conflicts of interest are defined as remuneration in any amount from commercial interests; grants; research support; consulting fees; salary; ownership interests [*e.g.*, stocks and stock options (excluding diversified mutual funds)]; honoraria and other payments for participation in speakers' bureaus, advisory boards, or boards of directors; and all other financial benefits. Completed forms are available through the Endocrine Society office.

The Endocrine Society provided the funding for this guideline; the task force received no funding or remuneration from commercial or other entities.

Commissioned Systematic Review

The task force commissioned two systematic reviews to support this guideline. The first one aimed to summarize the available evidence on the effect of sex steroid use in transgender individuals on lipids and cardiovascular outcomes. The review identified 29 eligible studies at moderate risk of bias. In transgender males (female to male), sex steroid therapy was associated with a statistically significant increase in serum triglycerides and low-density lipoprotein cholesterol levels. High-density lipoprotein cholesterol levels decreased significantly across all follow-up time periods. In transgender females (male to female), serum triglycerides were significantly higher without any changes in other parameters. Few myocardial infarction, stroke, venous thromboembolism (VTE), and death events were reported. These events were more frequent in transgender females. However, the quality of the evidence was low. The second review summarized the available evidence regarding the effect of sex steroids on bone health in transgender individuals and identified 13 studies. In transgender males, there was no statistically significant difference in the lumbar spine, femoral neck, or total hip BMD at 12 and 24 months compared with baseline values before initiating masculinizing hormone therapy. In transgender females, there was a statistically significant increase in lumbar spine BMD at 12 months and 24 months compared with baseline values before initiation of feminizing hormone therapy. There was minimal information on fracture rates. The quality of evidence was also low.

Introduction

Throughout recorded history (in the absence of an endocrine disorder) some men and women have experienced confusion and anguish resulting from rigid, forced conformity to sexual dimorphism. In modern history, there have been numerous ongoing biological, psychological, cultural, political, and sociological debates over various aspects of gender variance. The 20th century marked the emergence of a social awakening for men and women with the belief that they are "trapped" in the wrong body (3). Magnus Hirschfeld and Harry Benjamin, among others, pioneered the medical responses to those who sought relief from and a resolution to their profound discomfort. Although the term transsexual became widely known after Benjamin wrote "The Transsexual Phenomenon" (4), it was Hirschfeld who coined the term "transsexual" in 1923 to describe people who want to live a life that corresponds with their experienced gender vs their designated gender (5). Magnus Hirschfeld (6) and others (4, 7) have described other types of trans phenomena besides transsexualism. These early researchers proposed that the gender identity of these people was located somewhere along a unidimensional continuum. This continuum ranged from all male through "something in between" to all female. Yet such a classification does not take into account that people may have gender identities outside this continuum. For instance, some experience themselves as having both a male and female gender identity, whereas others completely renounce any gender classification (8, 9). There are also reports of individuals experiencing a continuous and rapid involuntary alternation between a male and female identity (10) or men who do not experience themselves as men but do not want to live as women (11, 12). In some countries, (*e.g.*, Nepal, Bangladesh, and Australia), these nonmale or nonfemale genders are officially recognized (13). Specific treatment protocols, however, have not yet been developed for these groups.

Instead of the term transsexualism, the current classification system of the American Psychiatric Association uses the term gender dysphoria in its diagnosis of persons who are not satisfied with their designated gender (14). The current version of the World Health Organization's ICD-10 still uses the term transsexualism when diagnosing adolescents and adults. However, for the ICD-11, the World Health Organization has proposed using the term "gender incongruence" (15).

Treating persons with GD/gender incongruence (15) was previously limited to relatively ineffective elixirs or creams. However, more effective endocrinology-based treatments became possible with the availability of testosterone in 1935 and diethylstilbestrol in 1938. Reports of individuals with GD/gender incongruence who were treated with hormones and gender-affirming surgery appeared in the press during the second half of the 20th century. The Harry Benjamin International Gender Dysphoria Association was founded in September 1979 and is now called the World Professional Association for Transgender Health (WPATH). WPATH published its first Standards of Care in 1979. These standards have since been regularly updated, providing guidance for treating persons with GD/gender incongruence (16).

Prior to 1975, few peer-reviewed articles were published concerning endocrine treatment of transgender persons. Since then, more than two thousand articles about various aspects of transgender care have appeared.

It is the purpose of this guideline to make detailed recommendations and suggestions, based on existing medical literature and clinical experience, that will enable treating physicians to maximize benefit and minimize risk when caring for individuals diagnosed with GD/gender incongruence.

In the future, we need more rigorous evaluations of the effectiveness and safety of endocrine and surgical protocols. Specifically, endocrine treatment protocols for GD/gender incongruence should include the careful assessment of the following: (1) the effects of prolonged delay of puberty in adolescents on bone health, gonadal function, and the brain (including effects on cognitive, emotional, social, and sexual development); (2) the effects of treatment in adults on sex hormone levels; (3) the requirement for and the effects of progestins and other agents used to suppress endogenous sex steroids during treatment; and (4) the risks and benefits of gender-affirming hormone treatment in older transgender people.

To successfully establish and enact these protocols, a commitment of mental health and endocrine investigators is required to collaborate in long-term, large-scale studies across countries that use the same diagnostic and inclusion criteria, medications, assay methods, and response assessment tools (*e.g.*, the European Network for the Investigation of Gender Incongruence) (17, 18).

Terminology and its use vary and continue to evolve. Table 1 contains the definitions of terms as they are used throughout this guideline.

Biological Determinants of Gender Identity Development

One's self-awareness as male or female changes gradually during infant life and childhood. This process of cognitive and affective learning evolves with interactions with parents, peers, and environment. A fairly accurate timetable exists outlining the steps in this process (19). Normative psychological literature, however, does not address if and when gender identity becomes crystallized and what factors contribute to the development of a gender identity that is not congruent with the gender of rearing. Results of studies from a variety of biomedical disciplines—genetic, endocrine, and neuroanatomic—support the concept that gender identity and/or gender expression (20) likely reflect a complex interplay of biological, environmental, and cultural factors (21, 22).

With respect to endocrine considerations, studies have failed to find differences in circulating levels of sex steroids between transgender and nontransgender individuals (23). However, studies in individuals with a disorder/difference of sex development (DSD) have informed our understanding of the role that hormones may play in gender identity outcome, even though most persons with GD/gender incongruence do not have a DSD. For example, although most 46,XX adult individuals with virilizing congenital adrenal hyperplasia caused by mutations in *CYP21A2* reported a female gender identity, the prevalence of GD/gender incongruence was much greater in this group than in the general population without a DSD. This supports the concept that there is a role for prenatal/postnatal androgens in gender development (24–26), although some studies indicate that prenatal androgens are more likely to affect gender behavior and sexual orientation rather than gender identity *per se* (27, 28).

Researchers have made similar observations regarding the potential role of androgens in the development of gender identity in other individuals with DSD. For example, a review of two groups of 46,XY persons, each with androgen synthesis deficiencies and female raised, reported transgender male (female-to-male) gender role changes in 56% to 63% and 39% to 64% of patients, respectively (29). Also, in 46,XY female-raised individuals with cloacal

doi: 10.1210/jc.2017-01658

Table 1. Definitions of Terms Used in This Guideline

Biological sex, biological male or female: These terms refer to physical aspects of maleness and femaleness. As these may not be in line with each other (*e.g.*, a person with XY chromosomes may have female-appearing genitalia), the terms biological sex and biological male or female are imprecise and should be avoided.
Cisgender: This means not transgender. An alternative way to describe individuals who are not transgender is "non-transgender people."
Gender-affirming (hormone) treatment: See "gender reassignment"
Gender dysphoria: This is the distress and unease experienced if gender identity and designated gender are not completely congruent (see Table 2). In 2013, the American Psychiatric Association released the fifth edition of the DSM-5, which replaced "gender identity disorder" with "gender dysphoria" and changed the criteria for diagnosis.
Gender expression: This refers to external manifestations of gender, expressed through one's name, pronouns, clothing, haircut, behavior, voice, or body characteristics. Typically, transgender people seek to make their gender expression align with their gender identity, rather than their designated gender.
Gender identity/experienced gender: This refers to one's internal, deeply held sense of gender. For transgender people, their gender identity does not match their sex designated at birth. Most people have a gender identity of man or woman (or boy or girl). For some people, their gender identity does not fit neatly into one of those two choices. Unlike gender expression (see below), gender identity is not visible to others.
Gender identity disorder: This is the term used for GD/gender incongruence in previous versions of DSM (see "gender dysphoria"). The ICD-10 still uses the term for diagnosing child diagnoses, but the upcoming ICD-11 has proposed using "gender incongruence of childhood."
Gender incongruence: This is an umbrella term used when the gender identity and/or gender expression differs from what is typically associated with the designated gender. Gender incongruence is also the proposed name of the gender identity–related diagnoses in ICD-11. Not all individuals with gender incongruence have gender dysphoria or seek treatment.
Gender variance: See "gender incongruence"
Gender reassignment: This refers to the treatment procedure for those who want to adapt their bodies to the experienced gender by means of hormones and/or surgery. This is also called gender-confirming or gender-affirming treatment.
Gender-reassignment surgery (gender-confirming/gender-affirming surgery): These terms refer only to the surgical part of gender-confirming/gender-affirming treatment.
Gender role: This refers to behaviors, attitudes, and personality traits that a society (in a given culture and historical period) designates as masculine or feminine and/or that society associates with or considers typical of the social role of men or women.
Sex designated at birth: This refers to sex assigned at birth, usually based on genital anatomy.
Sex: This refers to attributes that characterize biological maleness or femaleness. The best known attributes include the sex-determining genes, the sex chromosomes, the H-Y antigen, the gonads, sex hormones, internal and external genitalia, and secondary sex characteristics.
Sexual orientation: This term describes an individual's enduring physical and emotional attraction to another person. Gender identity and sexual orientation are not the same. Irrespective of their gender identity, transgender people may be attracted to women (gynephilic), attracted to men (androphilic), bisexual, asexual, or queer.
Transgender: This is an umbrella term for people whose gender identity and/or gender expression differs from what is typically associated with their sex designated at birth. Not all transgender individuals seek treatment.
Transgender male (also: trans man, female-to-male, transgender male): This refers to individuals assigned female at birth but who identify and live as men.
Transgender woman (also: trans woman, male-to female, transgender female): This refers to individuals assigned male at birth but who identify and live as women.
Transition: This refers to the process during which transgender persons change their physical, social, and/or legal characteristics consistent with the affirmed gender identity. Prepubertal children may choose to transition socially.
Transsexual: This is an older term that originated in the medical and psychological communities to refer to individuals who have permanently transitioned through medical interventions or desired to do so.

exstrophy and penile agenesis, the occurrence of transgender male changes was significantly more prevalent than in the general population (30, 31). However, the fact that a high percentage of individuals with the same conditions did not change gender suggests that cultural factors may play a role as well.

With respect to genetics and gender identity, several studies have suggested heritability of GD/gender incongruence (32, 33). In particular, a study by Heylens *et al.* (33) demonstrated a 39.1% concordance rate for gender identity disorder (based on the DSM-IV criteria) in 23 monozygotic twin pairs but no concordance in 21 same-sex dizygotic or seven opposite-sex twin pairs. Although numerous investigators have sought to identify specific genes associated with GD/gender incongruence, such studies have been inconsistent and without strong statistical significance (34–38).

Studies focusing on brain structure suggest that the brain phenotypes of people with GD/gender incongruence differ in various ways from control males and females, but that there is not a complete sex reversal in brain structures (39).

In summary, although there is much that is still unknown with respect to gender identity and its expression, compelling studies support the concept that biologic factors, in addition to environmental factors, contribute to this fundamental aspect of human development.

Natural History of Children With GD/Gender Incongruence

With current knowledge, we cannot predict the psychosexual outcome for any specific child. Prospective follow-up studies show that childhood GD/gender incongruence does not invariably persist into adolescence and adulthood (so-called "desisters"). Combining all outcome studies to date, the GD/gender incongruence of a minority of prepubertal children appears to persist in adolescence (20, 40). In adolescence, a significant number of these desisters identify as homosexual or bisexual. It may be that children who only showed some gender nonconforming characteristics have been included in the follow-up studies, because the DSM-IV text revision criteria for a diagnosis were rather broad. However, the persistence of GD/gender incongruence into adolescence is more likely if it had been extreme in childhood (41, 42). With the newer, stricter criteria of the DSM-5 (Table 2), persistence rates may well be different in future studies.

1.0 Evaluation of Youth and Adults

Gender-affirming treatment is a multidisciplinary effort. After evaluation, education, and diagnosis, treatment may include mental health care, hormone therapy, and/or surgical therapy. Together with an MHP, hormone-prescribing clinicians should examine the psychosocial impact of the potential changes on people's lives, including mental health, friends, family, jobs, and their role in society. Transgender individuals should be encouraged to experience living in the new gender role and assess whether this improves their quality of life. Although the focus of this guideline is gender-affirming hormone therapy, collaboration with appropriate professionals responsible for each aspect of treatment maximizes a successful outcome.

Diagnostic assessment and mental health care

GD/gender incongruence may be accompanied with psychological or psychiatric problems (43–51). It is therefore necessary that clinicians who prescribe hormones and are involved in diagnosis and psychosocial assessment meet the following criteria: (1) are competent in using the DSM and/or the ICD for diagnostic purposes, (2) are able to diagnose GD/gender incongruence and make a distinction between GD/gender incongruence and conditions that have similar features (*e.g.*, body dysmorphic disorder), (3) are trained in diagnosing psychiatric conditions, (4) undertake or refer for appropriate treatment, (5) are able to do a psychosocial assessment of the patient's understanding, mental health, and social conditions that can impact gender-affirming hormone therapy, and (6) regularly attend relevant professional meetings.

Because of the psychological vulnerability of many individuals with GD/gender incongruence, it is important that mental health care is available before, during, and sometimes also after transitioning. For children and adolescents, an MHP who has training/experience in child and adolescent gender development (as well as child and adolescent psychopathology) should make the diagnosis, because assessing GD/gender incongruence in children and adolescents is often extremely complex.

During assessment, the clinician obtains information from the individual seeking gender-affirming treatment. In the case

Table 2. DSM-5 Criteria for Gender Dysphoria in Adolescents and Adults

A. A marked incongruence between one's experienced/expressed gender and natal gender of at least 6 mo in duration, as manifested by at least two of the following:
   1. A marked incongruence between one's experienced/expressed gender and primary and/or secondary sex characteristics (or in young adolescents, the anticipated secondary sex characteristics)
   2. A strong desire to be rid of one's primary and/or secondary sex characteristics because of a marked incongruence with one's experienced/expressed gender (or in young adolescents, a desire to prevent the development of the anticipated secondary sex characteristics)
   3. A strong desire for the primary and/or secondary sex characteristics of the other gender
   4. A strong desire to be of the other gender (or some alternative gender different from one's designated gender)
   5. A strong desire to be treated as the other gender (or some alternative gender different from one's designated gender)
   6. A strong conviction that one has the typical feelings and reactions of the other gender (or some alternative gender different from one's designated gender)

B. The condition is associated with clinically significant distress or impairment in social, occupational, or other important areas of functioning.

Specify if:
   1. The condition exists with a disorder of sex development.
   2. The condition is posttransitional, in that the individual has transitioned to full-time living in the desired gender (with or without legalization of gender change) and has undergone (or is preparing to have) at least one sex-related medical procedure or treatment regimen—namely, regular sex hormone treatment or gender reassignment surgery confirming the desired gender (*e.g.*, penectomy, vaginoplasty in natal males; mastectomy or phalloplasty in natal females).

Reference: American Psychiatric Association (14).

doi: 10.1210/jc.2017-01658

of adolescents, the clinician also obtains information from the parents or guardians regarding various aspects of the child's general and psychosexual development and current functioning. On the basis of this information, the clinician:

- decides whether the individual fulfills criteria for treatment (see Tables 2 and 3) for GD/gender incongruence (DSM-5) or transsexualism (DSM-5 and/or ICD-10);
- informs the individual about the possibilities and limitations of various kinds of treatment (hormonal/surgical and nonhormonal), and if medical treatment is desired, provides correct information to prevent unrealistically high expectations;
- assesses whether medical interventions may result in unfavorable psychological and social outcomes.

In cases in which severe psychopathology, circumstances, or both seriously interfere with the diagnostic work or make satisfactory treatment unlikely, clinicians should assist the adolescent in managing these other issues. Literature on postoperative regret suggests that besides poor quality of surgery, severe psychiatric comorbidity and lack of support may interfere with positive outcomes (52–56).

For adolescents, the diagnostic procedure usually includes a complete psychodiagnostic assessment (57) and an assessment of the decision-making capability of the youth. An evaluation to assess the family's ability to endure stress, give support, and deal with the complexities of the adolescent's situation should be part of the diagnostic phase (58).

Social transitioning

A change in gender expression and role (which may involve living part time or full time in another gender role that is consistent with one's gender identity) may test the person's resolve, the capacity to function in the affirmed gender, and the adequacy of social, economic, and psychological supports. It assists both the individual and the clinician in their judgments about how to proceed (16). During social transitioning, the person's feelings about the social transformation (including coping with the responses of others) is a major focus of the counseling. The optimal timing for social transitioning may differ between individuals. Sometimes people wait until they start gender-affirming hormone treatment to make social transitioning easier, but individuals increasingly start social transitioning long before they receive medically supervised, gender-affirming hormone treatment.

Criteria

Adolescents and adults seeking gender-affirming hormone treatment and surgery should satisfy certain criteria before proceeding (16). Criteria for gender-affirming hormone therapy for adults are in Table 4, and criteria for gender-affirming hormone therapy for adolescents are in Table 5. Follow-up studies in adults meeting these criteria indicate a high satisfaction rate with treatment (59). However, the quality of evidence is usually low. A few follow-up studies on adolescents who fulfilled these criteria also indicated good treatment results (60–63).

Recommendations for Those Involved in the Gender-Affirming Hormone Treatment of Individuals With GD/Gender Incongruence

1.1. We advise that only trained MHPs who meet the following criteria should diagnose GD/gender incongruence in adults: (1) competence in using the DSM and/or the ICD for diagnostic purposes, (2) the ability to diagnose GD/gender incongruence and make a distinction between GD/gender incongruence and conditions that have similar features (*e.g.*, body dysmorphic disorder), (3) training in diagnosing psychiatric conditions, (4) the ability to undertake or refer for appropriate treatment, (5) the ability to psychosocially assess the person's understanding, mental health, and social conditions that can impact gender-affirming hormone therapy, and (6) a practice of regularly attending relevant professional meetings. (Ungraded Good Practice Statement)

1.2. We advise that only MHPs who meet the following criteria should diagnose GD/gender incongruence in children and adolescents: (1) training in child and adolescent developmental psychology and psychopathology, (2) competence in using the DSM and/or ICD for diagnostic

Table 3. ICD-10 Criteria for Transsexualism

Transsexualism (F64.0) has three criteria:
1. The desire to live and be accepted as a member of the opposite sex, usually accompanied by the wish to make his or her body as congruent as possible with the preferred sex through surgery and hormone treatments.
2. The transsexual identity has been present persistently for at least 2 y.
3. The disorder is not a symptom of another mental disorder or a genetic, DSD, or chromosomal abnormality.

Table 4. Criteria for Gender-Affirming Hormone Therapy for Adults

1. Persistent, well-documented gender dysphoria/gender incongruence
2. The capacity to make a fully informed decision and to consent for treatment
3. The age of majority in a given country (if younger, follow the criteria for adolescents)
4. Mental health concerns, if present, must be reasonably well controlled

Reproduced from World Professional Association for Transgender Health (16).

purposes, (3) the ability to make a distinction between GD/gender incongruence and conditions that have similar features (*e.g.*, body dysmorphic disorder), (4) training in diagnosing psychiatric conditions, (5) the ability to undertake or refer for appropriate treatment, (6) the ability to psychosocially assess the person's understanding and social conditions that can impact gender-affirming hormone therapy, (7) a practice of regularly attending relevant professional meetings, and (8) knowledge of the criteria for puberty blocking and gender-affirming hormone treatment in adolescents. (Ungraded Good Practice Statement)

Evidence

Individuals with gender identity issues may have psychological or psychiatric problems (43–48, 50, 51, 64, 65). It is therefore necessary that clinicians making the diagnosis are able to make a distinction between GD/gender incongruence and conditions that have similar features. Examples of conditions with similar features are body dysmorphic disorder, body identity integrity disorder (a condition in which individuals have a sense that their anatomical configuration as an able-bodied person is somehow wrong or inappropriate) (66), or certain forms of eunuchism (in which a person is preoccupied with or engages in castration and/or penectomy for

Table 5. Criteria for Gender-Affirming Hormone Therapy for Adolescents

Adolescents are eligible for GnRH agonist treatment if:

1. A qualified MHP has confirmed that:
- the adolescent has demonstrated a long-lasting and intense pattern of gender nonconformity or gender dysphoria (whether suppressed or expressed),
- gender dysphoria worsened with the onset of puberty,
- any coexisting psychological, medical, or social problems that could interfere with treatment (*e.g.*, that may compromise treatment adherence) have been addressed, such that the adolescent's situation and functioning are stable enough to start treatment,
- the adolescent has sufficient mental capacity to give informed consent to this (reversible) treatment,

2. And the adolescent:
- has been informed of the effects and side effects of treatment (including potential loss of fertility if the individual subsequently continues with sex hormone treatment) and options to preserve fertility,
- has given informed consent and (particularly when the adolescent has not reached the age of legal medical consent, depending on applicable legislation) the parents or other caretakers or guardians have consented to the treatment and are involved in supporting the adolescent throughout the treatment process,

3. And a pediatric endocrinologist or other clinician experienced in pubertal assessment
- agrees with the indication for GnRH agonist treatment,
- has confirmed that puberty has started in the adolescent (Tanner stage ≥G2/B2),
- has confirmed that there are no medical contraindications to GnRH agonist treatment.

Adolescents are eligible for subsequent sex hormone treatment if:

1. A qualified MHP has confirmed:
- the persistence of gender dysphoria,
- any coexisting psychological, medical, or social problems that could interfere with treatment (*e.g.*, that may compromise treatment adherence) have been addressed, such that the adolescent's situation and functioning are stable enough to start sex hormone treatment,
- the adolescent has sufficient mental capacity (which most adolescents have by age 16 years) to estimate the consequences of this (partly) irreversible treatment, weigh the benefits and risks, and give informed consent to this (partly) irreversible treatment,

2. And the adolescent:
- has been informed of the (irreversible) effects and side effects of treatment (including potential loss of fertility and options to preserve fertility),
- has given informed consent and (particularly when the adolescent has not reached the age of legal medical consent, depending on applicable legislation) the parents or other caretakers or guardians have consented to the treatment and are involved in supporting the adolescent throughout the treatment process,

3. And a pediatric endocrinologist or other clinician experienced in pubertal induction:
- agrees with the indication for sex hormone treatment,
- has confirmed that there are no medical contraindications to sex hormone treatment.

Reproduced from World Professional Association for Transgender Health (16).

doi: 10.1210/jc.2017-01658

reasons that are not gender identity related) (11). Clinicians should also be able to diagnose psychiatric conditions accurately and ensure that these conditions are treated appropriately, particularly when the conditions may complicate treatment, affect the outcome of gender-affirming treatment, or be affected by hormone use.

Values and preferences

The task force placed a very high value on avoiding harm from hormone treatment in individuals who have conditions other than GD/gender incongruence and who may not benefit from the physical changes associated with this treatment and placed a low value on any potential benefit these persons believe they may derive from hormone treatment. This justifies the good practice statement.

1.3. We advise that decisions regarding the social transition of prepubertal youths with GD/gender incongruence are made with the assistance of an MHP or another experienced professional. (Ungraded Good Practice Statement).

1.4. We recommend against puberty blocking and gender-affirming hormone treatment in prepubertal children with GD/gender incongruence. (1 |⊕⊕◯◯)

Evidence

In most children diagnosed with GD/gender incongruence, it did not persist into adolescence. The percentages differed among studies, probably dependent on which version of the DSM clinicians used, the patient's age, the recruitment criteria, and perhaps cultural factors. However, the large majority (about 85%) of prepubertal children with a childhood diagnosis did not remain GD/gender incongruent in adolescence (20). If children have completely socially transitioned, they may have great difficulty in returning to the original gender role upon entering puberty (40). Social transition is associated with the persistence of GD/gender incongruence as a child progresses into adolescence. It may be that the presence of GD/gender incongruence in prepubertal children is the earliest sign that a child is destined to be transgender as an adolescent/adult (20). However, social transition (in addition to GD/gender incongruence) has been found to contribute to the likelihood of persistence.

This recommendation, however, does not imply that children should be discouraged from showing gender-variant behaviors or should be punished for exhibiting such behaviors. In individual cases, an early complete social transition may result in a more favorable outcome, but there are currently no criteria to identify the GD/gender-incongruent children to whom this applies. At the present time, clinical experience suggests that persistence of GD/gender incongruence can only be reliably assessed after the first signs of puberty.

Values and preferences

The task force placed a high value on avoiding harm with gender-affirming hormone therapy in prepubertal children with GD/gender incongruence. This justifies the strong recommendation in the face of low-quality evidence.

1.5. We recommend that clinicians inform and counsel all individuals seeking gender-affirming medical treatment regarding options for fertility preservation prior to initiating puberty suppression in adolescents and prior to treating with hormonal therapy of the affirmed gender in both adolescents and adults. (1 |⊕⊕⊕◯)

Remarks

Persons considering hormone use for gender affirmation need adequate information about this treatment in general and about fertility effects of hormone treatment in particular to make an informed and balanced decision (67, 68). Because young adolescents may not feel qualified to make decisions about fertility and may not fully understand the potential effects of hormonal interventions, consent and protocol education should include parents, the referring MHP(s), and other members of the adolescent's support group. To our knowledge, there are no formally evaluated decision aids available to assist in the discussion and decision regarding the future fertility of adolescents or adults beginning gender-affirming treatment.

Treating early pubertal youth with GnRH analogs will temporarily impair spermatogenesis and oocyte maturation. Given that an increasing number of transgender youth want to preserve fertility potential, delaying or temporarily discontinuing GnRH analogs to promote gamete maturation is an option. This option is often not preferred, because mature sperm production is associated with later stages of puberty and with the significant development of secondary sex characteristics.

For those designated male at birth with GD/gender incongruence and who are in early puberty, sperm production and the development of the reproductive tract are insufficient for the cryopreservation of sperm. However, prolonged pubertal suppression using GnRH analogs is reversible and clinicians should inform these individuals that sperm production can be initiated following prolonged gonadotropin suppression. This can be accomplished by spontaneous gonadotropin recovery after

cessation of GnRH analogs or by gonadotropin treatment and will probably be associated with physical manifestations of testosterone production, as stated above. Note that there are no data in this population concerning the time required for sufficient spermatogenesis to collect enough sperm for later fertility. In males treated for precocious puberty, spermarche was reported 0.7 to 3 years after cessation of GnRH analogs (69). In adult men with gonadotropin deficiency, sperm are noted in seminal fluid by 6 to 12 months of gonadotropin treatment. However, sperm numbers when partners of these patients conceive are far below the "normal range" (70, 71).

In girls, no studies have reported long-term, adverse effects of pubertal suppression on ovarian function after treatment cessation (72, 73). Clinicians should inform adolescents that no data are available regarding either time to spontaneous ovulation after cessation of GnRH analogs or the response to ovulation induction following prolonged gonadotropin suppression.

In males with GD/gender incongruence, when medical treatment is started in a later phase of puberty or in adulthood, spermatogenesis is sufficient for cryopreservation and storage of sperm. *In vitro* spermatogenesis is currently under investigation. Restoration of spermatogenesis after prolonged estrogen treatment has not been studied.

In females with GD/gender incongruence, the effect of prolonged treatment with exogenous testosterone on ovarian function is uncertain. There have been reports of an increased incidence of polycystic ovaries in transgender males, both prior to and as a result of androgen treatment (74–77), although these reports were not confirmed by others (78). Pregnancy has been reported in transgender males who have had prolonged androgen treatment and have discontinued testosterone but have not had genital surgery (79, 80). A reproductive endocrine gynecologist can counsel patients before gender-affirming hormone treatment or surgery regarding potential fertility options (81). Techniques for cryopreservation of oocytes, embryos, and ovarian tissue continue to improve, and oocyte maturation of immature tissue is being studied (82).

2.0 Treatment of Adolescents

During the past decade, clinicians have progressively acknowledged the suffering of young adolescents with GD/gender incongruence. In some forms of GD/gender incongruence, psychological interventions may be useful and sufficient. However, for many adolescents with GD/gender incongruence, the pubertal physical changes are unbearable. As early medical intervention may prevent psychological harm, various clinics have decided to start treating young adolescents with GD/gender incongruence with puberty-suppressing medication (a GnRH analog). As compared with starting gender-affirming treatment long after the first phases of puberty, a benefit of pubertal suppression at early puberty may be a better psychological and physical outcome.

In girls, the first physical sign of puberty is the budding of the breasts followed by an increase in breast and fat tissue. Breast development is also associated with the pubertal growth spurt, and menarche occurs ~2 years later. In boys, the first physical change is testicular growth. A testicular volume ≥4 mL is seen as consistent with the initiation of physical puberty. At the beginning of puberty, estradiol and testosterone levels are still low and are best measured in the early morning with an ultrasensitive assay. From a testicular volume of 10 mL, daytime testosterone levels increase, leading to virilization (83). Note that pubic hair and/or axillary hair/odor may not reflect the onset of gonadarche; instead, it may reflect adrenarche alone.

2.1. We suggest that adolescents who meet diagnostic criteria for GD/gender incongruence, fulfill criteria for treatment (Table 5), and are requesting treatment should initially undergo treatment to suppress pubertal development. (2 ⊕⊕◯◯)

2.2. We suggest that clinicians begin pubertal hormone suppression after girls and boys first exhibit physical changes of puberty (Tanner stages G2/B2). (2 ⊕⊕◯◯)

Evidence

Pubertal suppression can expand the diagnostic phase by a long period, giving the subject more time to explore options and to live in the experienced gender before making a decision to proceed with gender-affirming sex hormone treatments and/or surgery, some of which is irreversible (84, 85). Pubertal suppression is fully reversible, enabling full pubertal development in the natal gender, after cessation of treatment, if appropriate. The experience of full endogenous puberty is an undesirable condition for the GD/gender-incongruent individual and may seriously interfere with healthy psychological functioning and well-being. Treating GD/gender-incongruent adolescents entering puberty with GnRH analogs has been shown to improve psychological functioning in several domains (86).

Another reason to start blocking pubertal hormones early in puberty is that the physical outcome is improved compared with initiating physical transition after puberty has been completed (60, 62). Looking like a man or woman when living as the opposite sex creates difficult

doi: 10.1210/jc.2017-01658

barriers with enormous life-long disadvantages. We therefore advise starting suppression in early puberty to prevent the irreversible development of undesirable secondary sex characteristics. However, adolescents with GD/gender incongruence should experience the first changes of their endogenous spontaneous puberty, because their emotional reaction to these first physical changes has diagnostic value in establishing the persistence of GD/gender incongruence (85). Thus, Tanner stage 2 is the optimal time to start pubertal suppression. However, pubertal suppression treatment in early puberty will limit the growth of the penis and scrotum, which will have a potential effect on future surgical treatments (87).

Clinicians can also use pubertal suppression in adolescents in later pubertal stages to stop menses in transgender males and prevent facial hair growth in transgender females. However, in contrast to the effects in early pubertal adolescents, physical sex characteristics (such as more advanced breast development in transgender boys and lowering of the voice and outgrowth of the jaw and brow in transgender girls) are not reversible.

Values and preferences

These recommendations place a high value on avoiding an unsatisfactory physical outcome when secondary sex characteristics have become manifest and irreversible, a higher value on psychological well-being, and a lower value on avoiding potential harm from early pubertal suppression.

Remarks

Table 6 lists the Tanner stages of breast and male genital development. Careful documentation of hallmarks of pubertal development will ensure precise timing when initiating pubertal suppression once puberty has started. Clinicians can use pubertal LH and sex steroid levels to confirm that puberty has progressed sufficiently before starting pubertal suppression (88). Reference ranges for sex steroids by Tanner stage may vary depending on the assay used. Ultrasensitive sex steroid and gonadotropin assays will help clinicians document early pubertal changes.

Irreversible and, for GD/gender-incongruent adolescents, undesirable sex characteristics in female puberty are breasts, female body habitus, and, in some cases, relative short stature. In male puberty, they are a prominent Adam's apple; low voice; male bone configuration, such as a large jaw, big feet and hands, and tall stature; and male hair pattern on the face and extremities.

2.3. We recommend that, where indicated, GnRH analogues are used to suppress pubertal hormones. (1 |⊕⊕◯◯)

Evidence

Clinicians can suppress pubertal development and gonadal function most effectively via gonadotropin suppression using GnRH analogs. GnRH analogs are long-acting agonists that suppress gonadotropins by GnRH receptor desensitization after an initial increase of gonadotropins during ~10 days after the first and (to a lesser degree) the second injection (89). Antagonists immediately suppress pituitary gonadotropin secretion (90, 91). Long-acting GnRH analogs are the currently preferred treatment option. Clinicians may consider long-acting GnRH antagonists when evidence on their safety and efficacy in adolescents becomes available.

During GnRH analog treatment, slight development of secondary sex characteristics may regress, and in a later phase of pubertal development, it will stop. In girls, breast tissue will become atrophic, and menses will stop. In boys, virilization will stop, and testicular volume may decrease (92).

An advantage of using GnRH analogs is the reversibility of the intervention. If, after extensive exploration of his/her transition wish, the individual no longer desires transition, they can discontinue pubertal suppression. In subjects with

Table 6. Tanner Stages of Breast Development and Male External Genitalia

The description of Tanner stages for breast development:
1. Prepubertal
2. Breast and papilla elevated as small mound; areolar diameter increased
3. Breast and areola enlarged, no contour separation
4. Areola and papilla form secondary mound
5. Mature; nipple projects, areola part of general breast contour

For penis and testes:
1. Prepubertal, testicular volume <4 mL
2. Slight enlargement of penis; enlarged scrotum, pink, texture altered, testes 4–6 mL
3. Penis longer, testes larger (8–12 mL)
4. Penis and glans larger, including increase in breadth; testes larger (12–15 mL), scrotum dark
5. Penis adult size; testicular volume > 15 ml

Adapted from Lawrence (56).

precocious puberty, spontaneous pubertal development has been shown to resume after patients discontinue taking GnRH analogs (93).

Recommendations 2.1 to 2.3 are supported by a prospective follow-up study from The Netherlands. This report assessed mental health outcomes in 55 transgender adolescents/young adults (22 transgender females and 33 transgender males) at three time points: (1) before the start of GnRH agonist (average age of 14.8 years at start of treatment), (2) at initiation of gender-affirming hormones (average age of 16.7 years at start of treatment), and (3) 1 year after "gender-reassignment surgery" (average age of 20.7 years) (63). Despite a decrease in depression and an improvement in general mental health functioning, GD/gender incongruence persisted through pubertal suppression, as previously reported (86). However, following sex hormone treatment and gender-reassignment surgery, GD/gender incongruence was resolved and psychological functioning steadily improved (63). Furthermore, well-being was similar to or better than that reported by age-matched young adults from the general population, and none of the study participants regretted treatment. This study represents the first long-term follow-up of individuals managed according to currently existing clinical practice guidelines for transgender youth, and it underscores the benefit of the multidisciplinary approach pioneered in The Netherlands; however, further studies are needed.

Side effects

The primary risks of pubertal suppression in GD/gender-incongruent adolescents may include adverse effects on bone mineralization (which can theoretically be reversed with sex hormone treatment), compromised fertility if the person subsequently is treated with sex hormones, and unknown effects on brain development. Few data are available on the effect of GnRH analogs on BMD in adolescents with GD/gender incongruence. Initial data in GD/gender-incongruent subjects demonstrated no change of absolute areal BMD during 2 years of GnRH analog therapy but a decrease in BMD *z* scores (85). A recent study also suggested suboptimal bone mineral accrual during GnRH analog treatment. The study reported a decrease in areal BMD *z* scores and of bone mineral apparent density *z* scores (which takes the size of the bone into account) in 19 transgender males treated with GnRH analogs from a mean age of 15.0 years (standard deviation = 2.0 years) for a median duration of 1.5 years (0.3 to 5.2 years) and in 15 transgender females treated from 14.9 (±1.9) years for 1.3 years (0.5 to 3.8 years), although not all changes were statistically significant (94). There was incomplete catch-up at age 22 years after sex hormone treatment from age 16.6 (±1.4) years for a median duration of 5.8 years (3.0 to 8.0 years) in transgender females and from age 16.4 (±2.3) years for 5.4 years (2.8 to 7.8 years) in transgender males. Little is known about more prolonged use of GnRH analogs. Researchers reported normal BMD *z* scores at age 35 years in one individual who used GnRH analogs from age 13.7 years until age 18.6 years before initiating sex hormone treatment (65).

Additional data are available from individuals with late puberty or GnRH analog treatment of other indications. Some studies reported that men with constitutionally delayed puberty have decreased BMD in adulthood (95). However, other studies reported that these men have normal BMD (96, 97). Treating adults with GnRH analogs results in a decrease of BMD (98). In children with central precocious puberty, treatment with GnRH analogs has been found to result in a decrease of BMD during treatment by some (99) but not others (100). Studies have reported normal BMD after discontinuing therapy (69, 72, 73, 101, 102). In adolescents treated with growth hormone who are small for gestational age and have normal pubertal timing, 2-year GnRH analog treatments did not adversely affect BMD (103). Calcium supplementation may be beneficial in optimizing bone health in GnRH analog–treated individuals (104). There are no studies of vitamin D supplementation in this context, but clinicians should offer supplements to vitamin D–deficient adolescents. Physical activity, especially during growth, is important for bone mass in healthy individuals (103) and is therefore likely to be beneficial for bone health in GnRH analog–treated subjects.

GnRH analogs did not induce a change in body mass index standard deviation score in GD/gender-incongruent adolescents (94) but caused an increase in fat mass and decrease in lean body mass percentage (92). Studies in girls treated for precocious puberty also reported a stable body mass index standard deviation score during treatment (72) and body mass index and body composition comparable to controls after treatment (73).

Arterial hypertension has been reported as an adverse effect in a few girls treated with GnRH analogs for precocious/early puberty (105, 106). Blood pressure monitoring before and during treatment is recommended.

Individuals may also experience hot flashes, fatigue, and mood alterations as a consequence of pubertal suppression. There is no consensus on treatment of these side effects in this context.

It is recommended that any use of pubertal blockers (and subsequent use of sex hormones, as detailed below) include a discussion about implications for fertility (see recommendation 1.3). Transgender adolescents may

doi: 10.1210/jc.2017-01658

want to preserve fertility, which may be otherwise compromised if puberty is suppressed at an early stage and the individual completes phenotypic transition with the use of sex hormones.

Limited data are available regarding the effects of GnRH analogs on brain development. A single cross-sectional study demonstrated no compromise of executive function (107), but animal data suggest there may be an effect of GnRH analogs on cognitive function (108).

Values and preferences

Our recommendation of GnRH analogs places a higher value on the superior efficacy, safety, and reversibility of the pubertal hormone suppression achieved (as compared with the alternatives) and a relatively lower value on limiting the cost of therapy. Of the available alternatives, depot and oral progestin preparations are effective. Experience with this treatment dates back prior to the emergence of GnRH analogs for treating precocious puberty in papers from the 1960s and early 1970s (109–112). These compounds are usually safe, but some side effects have been reported (113–115). Only two recent studies involved transgender youth (116, 117). One of these studies described the use of oral lynestrenol monotherapy followed by the addition of testosterone treatment in transgender boys who were at Tanner stage B4 or further at the start of treatment (117). They found lynestrenol safe, but gonadotropins were not fully suppressed. The study reported metrorrhagia in approximately half of the individuals, mainly in the first 6 months. Acne, headache, hot flashes, and fatigue were other frequent side effects. Another progestin that has been studied in the United States is medroxyprogesterone. This agent is not as effective as GnRH analogs in lowering endogenous sex hormones either and may be associated with other side effects (116). Progestin preparations may be an acceptable treatment for persons without access to GnRH analogs or with a needle phobia. If GnRH analog treatment is not available (insurance denial, prohibitive cost, or other reasons), postpubertal, transgender female adolescents may be treated with an antiandrogen that directly suppresses androgen synthesis or action (see adult section).

Remarks

Measurements of gonadotropin and sex steroid levels give precise information about gonadal axis suppression, although there is insufficient evidence for any specific short-term monitoring scheme in children treated with GnRH analogs (88). If the gonadal axis is not completely suppressed—as evidenced by (for example) menses, erections, or progressive hair growth—the interval of GnRH analog treatment can be shortened or the dose increased. During treatment, adolescents should be monitored for negative effects of delaying puberty, including a halted growth spurt and impaired bone mineral accretion. Table 7 illustrates a suggested clinical protocol.

Anthropometric measurements and X-rays of the left hand to monitor bone age are informative for evaluating growth. To assess BMD, clinicians can perform dual-energy X-ray absorptiometry scans.

2.4. In adolescents who request sex hormone treatment (given this is a partly irreversible treatment), we recommend initiating treatment using a gradually increasing dose schedule (see Table 8) after a multidisciplinary team of medical and MHPs has confirmed the persistence of GD/gender incongruence and sufficient mental capacity to give informed consent, which most adolescents have by age 16 years (Table 5). (1 |⊕⊕◯◯)

2.5. We recognize that there may be compelling reasons to initiate sex hormone treatment prior to the age of 16 years in some adolescents with GD/gender incongruence, even though there are minimal published studies of gender-affirming hormone treatments administered before age 13.5 to 14 years. As with the care of adolescents ≥16 years of age, we recommend that an expert multidisciplinary team of medical and MHPs manage this treatment. (1 |⊕◯◯◯)

2.6. We suggest monitoring clinical pubertal development every 3 to 6 months and laboratory parameters every 6 to 12 months during sex hormone treatment (Table 9). (2 |⊕⊕◯◯)

Table 7. Baseline and Follow-Up Protocol During Suppression of Puberty

Every 3–6 mo
Anthropometry: height, weight, sitting height, blood pressure, Tanner stages
Every 6–12 mo
Laboratory: LH, FSH, E2/T, 25OH vitamin D
Every 1–2 y
Bone density using DXA
Bone age on X-ray of the left hand (if clinically indicated)

Adapted from Hembree *et al.* (118).

Abbreviations: DXA, dual-energy X-ray absorptiometry; E2, estradiol; FSH, follicle stimulating hormone; LH, luteinizing hormone; T, testosterone;

Table 8. Protocol Induction of Puberty

Induction of female puberty with oral 17β-estradiol, increasing the dose every 6 mo:
- 5 μg/kg/d
- 10 μg/kg/d
- 15 μg/kg/d
- 20 μg/kg/d
- Adult dose = 2–6 mg/d
- *In postpubertal transgender female adolescents, the dose of 17β-estradiol can be increased more rapidly:*
  - 1 mg/d for 6 mo
  - 2 mg/d

Induction of female puberty with transdermal 17β-estradiol, increasing the dose every 6 mo (new patch is placed every 3.5 d):
- 6.25–12.5 μg/24 h (cut 25-μg patch into quarters, then halves)
- 25 μg/24 h
- 37.5 μg/24 h
- Adult dose = 50–200 μg/24 h
- *For alternatives once at adult dose, see Table 11.*
- *Adjust maintenance dose to mimic physiological estradiol levels (see Table 15).*

Induction of male puberty with testosterone esters increasing the dose every 6 mo (IM or SC):
- 25 mg/m^2/2 wk (or alternatively, half this dose weekly, or double the dose every 4 wk)
- 50 mg/m^2/2 wk
- 75 mg/m^2/2 wk
- 100 mg/m^2/2 wk
- Adult dose = 100–200 mg every 2 wk
- *In postpubertal transgender male adolescents the dose of testosterone esters can be increased more rapidly:*
  - 75 mg/2 wk for 6 mo
  - 125 mg/2 wk
- *For alternatives once at adult dose, see Table 11.*
- *Adjust maintenance dose to mimic physiological testosterone levels (see Table 14).*

Adapted from Hembree *et al.* (118).

Abbreviations: IM, intramuscularly; SC, subcutaneously.

Evidence

Adolescents develop competence in decision making at their own pace. Ideally, the supervising medical professionals should individually assess this competence, although no objective tools to make such an assessment are currently available.

Many adolescents have achieved a reasonable level of competence by age 15 to 16 years (119), and in many countries 16-year-olds are legally competent with regard to medical decision making (120). However, others believe that although some capacities are generally achieved before age 16 years, other abilities (such as good risk assessment) do not develop until well after 18 years (121). They suggest that health care procedures should be divided along a matrix of relative risk, so that younger adolescents can be allowed to decide about low-risk procedures, such as most diagnostic tests and common therapies, but not about high-risk procedures, such as most surgical procedures (121).

Currently available data from transgender adolescents support treatment with sex hormones starting at age 16 years (63, 122). However, some patients may incur potential risks by waiting until age 16 years. These include the potential risk to bone health if puberty is suppressed

Table 9. Baseline and Follow-up Protocol During Induction of Puberty

Every 3–6 mo
- •Anthropometry: height, weight, sitting height, blood pressure, Tanner stages

Every 6–12 mo
- •In transgender males: hemoglobin/hematocrit, lipids, testosterone, 25OH vitamin D
- •In transgender females: prolactin, estradiol, 25OH vitamin D

Every 1–2 y
- •BMD using DXA
- •Bone age on X-ray of the left hand (if clinically indicated)

BMD should be monitored into adulthood (until the age of 25–30 y or until peak bone mass has been reached).
For recommendations on monitoring once pubertal induction has been completed, see Tables 14 and 15.

Adapted from Hembree *et al.* (118).

Abbreviation: DXA, dual-energy X-ray absorptiometry.

doi: 10.1210/jc.2017-01658

for 6 to 7 years before initiating sex hormones (*e.g.*, if someone reached Tanner stage 2 at age 9-10 years old). Additionally, there may be concerns about inappropriate height and potential harm to mental health (emotional and social isolation) if initiation of secondary sex characteristics must wait until the person has reached 16 years of age. However, only minimal data supporting earlier use of gender-affirming hormones in transgender adolescents currently exist (63). Clearly, long-term studies are needed to determine the optimal age of sex hormone treatment in GD/gender-incongruent adolescents.

The MHP who has followed the adolescent during GnRH analog treatment plays an essential role in assessing whether the adolescent is eligible to start sex hormone therapy and capable of consenting to this treatment (Table 5). Support of the family/environment is essential. Prior to the start of sex hormones, clinicians should discuss the implications for fertility (see recommendation 1.5). Throughout pubertal induction, an MHP and a pediatric endocrinologist (or other clinician competent in the evaluation and induction of pubertal development) should monitor the adolescent. In addition to monitoring therapy, it is also important to pay attention to general adolescent health issues, including healthy life style choices, such as not smoking, contraception, and appropriate vaccinations (*e.g.*, human papillomavirus).

For the induction of puberty, clinicians can use a similar dose scheme for hypogonadal adolescents with GD/gender incongruence as they use in other individuals with hypogonadism, carefully monitoring for desired and undesired effects (Table 8). In transgender female adolescents, transdermal 17β-estradiol may be an alternative for oral 17β-estradiol. It is increasingly used for pubertal induction in hypogonadal females. However, the absence of low-dose estrogen patches may be a problem. As a result, individuals may need to cut patches to size themselves to achieve appropriate dosing (123). In transgender male adolescents, clinicians can give testosterone injections intramuscularly or subcutaneously (124, 125).

When puberty is initiated with a gradually increasing schedule of sex steroid doses, the initial levels will not be high enough to suppress endogenous sex steroid secretion. Gonadotropin secretion and endogenous production of testosterone may resume and interfere with the effectiveness of estrogen treatment, in transgender female adolescents (126, 127). Therefore, continuation of GnRH analog treatment is advised until gonadectomy. Given that GD/gender-incongruent adolescents may opt not to have gonadectomy, long-term studies are necessary to examine the potential risks of prolonged GnRH analog treatment. Alternatively, in transgender male adolescents, GnRH analog treatment can be discontinued once an adult dose of testosterone has been reached and the individual is well virilized. If uterine bleeding occurs, a progestin can be added. However, the combined use of a GnRH analog (for ovarian suppression) and testosterone may enable phenotypic transition with a lower dose of testosterone in comparison with testosterone alone. If there is a wish or need to discontinue GnRH analog treatment in transgender female adolescents, they may be treated with an antiandrogen that directly suppresses androgen synthesis or action (see section 3.0 "Hormonal Therapy for Transgender Adults").

Values and preferences

The recommendation to initiate pubertal induction only when the individual has sufficient mental capacity (roughly age 16 years) to give informed consent for this partly irreversible treatment places a higher value on the ability of the adolescent to fully understand and oversee the partially irreversible consequences of sex hormone treatment and to give informed consent. It places a lower value on the possible negative effects of delayed puberty. We may not currently have the means to weigh adequately the potential benefits of waiting until around age 16 years to initiate sex hormones vs the potential risks/harm to BMD and the sense of social isolation from having the timing of puberty be so out of sync with peers (128).

Remarks

Before starting sex hormone treatment, effects on fertility and options for fertility preservation should be discussed. Adult height may be a concern in transgender adolescents. In a transgender female adolescent, clinicians may consider higher doses of estrogen or a more rapid tempo of dose escalation during pubertal induction. There are no established treatments yet to augment adult height in a transgender male adolescent with open epiphyses during pubertal induction. It is not uncommon for transgender adolescents to present for clinical services after having completed or nearly completed puberty. In such cases, induction of puberty with sex hormones can be done more rapidly (see Table 8). Additionally, an adult dose of testosterone in transgender male adolescents may suffice to suppress the gonadal axis without the need to use a separate agent. At the appropriate time, the multidisciplinary team should adequately prepare the adolescent for transition to adult care.

3.0 Hormonal Therapy for Transgender Adults

The two major goals of hormonal therapy are (1) to reduce endogenous sex hormone levels and thus reduce

the secondary sex characteristics of the individual's designated gender, and (2) to replace endogenous sex hormone levels consistent with the individual's gender identity by using the principles of hormone replacement treatment of hypogonadal patients. The timing of these two goals and the age at which to begin treatment with the sex hormones of the chosen gender is codetermined in collaboration with both the person pursuing transition and the health care providers. The treatment team should include a medical provider knowledgeable in transgender hormone therapy, an MHP knowledgeable in GD/gender incongruence and the mental health concerns of transition, and a primary care provider able to provide care appropriate for transgender individuals. The physical changes induced by this sex hormone transition are usually accompanied by an improvement in mental well-being (129, 130).

3.1. We recommend that clinicians confirm the diagnostic criteria of GD/gender incongruence and the criteria for the endocrine phase of gender transition before beginning treatment. (1 |⊕⊕⊕◯)

3.2. We recommend that clinicians evaluate and address medical conditions that can be exacerbated by hormone depletion and treatment with sex hormones of the affirmed gender before beginning treatment (Table 10). (1 |⊕⊕⊕◯)

3.3. We suggest that clinicians measure hormone levels during treatment to ensure that endogenous sex steroids are suppressed and administered sex steroids are maintained in the normal physiologic range for the affirmed gender. (2 |⊕⊕◯◯)

Evidence

It is the responsibility of the treating clinician to confirm that the person fulfills criteria for treatment. The treating clinician should become familiar with the terms and criteria presented in Tables 1–5 and take a thorough history from the patient in collaboration with the other members of the treatment team. The treating clinician must ensure that the desire for transition is appropriate; the consequences, risks, and benefits of treatment are well understood; and the desire for transition persists. They also need to discuss fertility preservation options (see recommendation 1.3) (67, 68).

Transgender males

Clinical studies have demonstrated the efficacy of several different androgen preparations to induce masculinization in transgender males (Appendix A) (113, 114, 131–134). Regimens to change secondary sex characteristics follow the general principle of hormone replacement treatment of male hypogonadism (135). Clinicians can use either parenteral or transdermal preparations to achieve testosterone values in the normal male range (this is dependent on the specific assay, but is typically 320 to 1000 ng/dL) (Table 11) (136). Sustained supraphysiologic levels of testosterone increase the risk of adverse reactions (see section 4.0 "Adverse Outcome Prevention and Long-Term Care") and should be avoided.

Similar to androgen therapy in hypogonadal men, testosterone treatment in transgender males results in increased muscle mass and decreased fat mass, increased facial hair and acne, male pattern baldness in those genetically predisposed, and increased sexual desire (137).

Table 10. Medical Risks Associated With Sex Hormone Therapy

Transgender female: estrogen
- Very high risk of adverse outcomes:
  - •Thromboembolic disease
- Moderate risk of adverse outcomes:
  - •Macroprolactinoma
  - •Breast cancer
  - •Coronary artery disease
  - •Cerebrovascular disease
  - •Cholelithiasis
  - •Hypertriglyceridemia

Transgender male: testosterone
- Very high risk of adverse outcomes:
  - •Erythrocytosis (hematocrit > 50%)
- Moderate risk of adverse outcomes:
  - •Severe liver dysfunction (transaminases > threefold upper limit of normal)
  - •Coronary artery disease
  - •Cerebrovascular disease
  - •Hypertension
  - •Breast or uterine cancer

Table 11. Hormone Regimens in Transgender Persons

Transgender females[a]	
Estrogen	
Oral	
Estradiol	2.0–6.0 mg/d
Transdermal	
Estradiol transdermal patch (New patch placed every 3–5 d)	0.025–0.2 mg/d
Parenteral	
Estradiol valerate or cypionate	5–30 mg IM every 2 wk 2–10 mg IM every week
Anti-androgens	
Spironolactone	100–300 mg/d
Cyproterone acetate[b]	25–50 mg/d
GnRH agonist	3.75 mg SQ (SC) monthly 11.25 mg SQ (SC) 3-monthly
Transgender males	
Testosterone	
Parenteral testosterone	
Testosterone enanthate or cypionate	100–200 mg SQ (IM) every 2 wk or SQ (SC) 50% per week
Testosterone undecanoate[c]	1000 mg every 12 wk
Transdermal testosterone	
Testosterone gel 1.6%[d]	50–100 mg/d
Testosterone transdermal patch	2.5–7.5 mg/d

Abbreviations: IM, intramuscularly; SQ, sequentially; SC, subcutaneously.

[a]Estrogens used with or without antiandrogens or GnRH agonist.

[b]Not available in the United States.

[c]One thousand milligrams initially followed by an injection at 6 wk then at 12-wk intervals.

[d]Avoid cutaneous transfer to other individuals.

In transgender males, testosterone will result in clitoromegaly, temporary or permanent decreased fertility, deepening of the voice, cessation of menses (usually), and a significant increase in body hair, particularly on the face, chest, and abdomen. Cessation of menses may occur within a few months with testosterone treatment alone, although high doses of testosterone may be required. If uterine bleeding continues, clinicians may consider the addition of a progestational agent or endometrial ablation (138). Clinicians may also administer GnRH analogs or depot medroxyprogesterone to stop menses prior to testosterone treatment.

Transgender females

The hormone regimen for transgender females is more complex than the transgender male regimen (Appendix B). Treatment with physiologic doses of estrogen alone is insufficient to suppress testosterone levels into the normal range for females (139). Most published clinical studies report the need for adjunctive therapy to achieve testosterone levels in the female range (21, 113, 114, 132–134, 139, 140).

Multiple adjunctive medications are available, such as progestins with antiandrogen activity and GnRH agonists (141). Spironolactone works by directly blocking androgens during their interaction with the androgen receptor (114, 133, 142). It may also have estrogenic activity (143). Cyproterone acetate, a progestational compound with antiandrogenic properties (113, 132, 144), is widely used in Europe. 5α-Reductase inhibitors do not reduce testosterone levels and have adverse effects (145).

Dittrich *et al.* (141) reported that monthly doses of the GnRH agonist goserelin acetate in combination with estrogen were effective in reducing testosterone levels with a low incidence of adverse reactions in 60 transgender females. Leuprolide and transdermal estrogen were as effective as cyproterone and transdermal estrogen in a comparative retrospective study (146).

Patients can take estrogen as oral conjugated estrogens, oral 17β-estradiol, or transdermal 17β-estradiol. Among estrogen options, the increased risk of thromboembolic events associated with estrogens in general seems most concerning with ethinyl estradiol specifically (134, 140, 141), which is why we specifically suggest that it not be used in any transgender treatment plan. Data distinguishing among other estrogen options are less well established although there is some thought that oral routes of administration are more thrombogenic due to the "first pass effect" than are transdermal and parenteral routes, and that the risk of thromboembolic events is dose-dependent. Injectable estrogen and sublingual

estrogen may benefit from avoiding the first pass effect, but they can result in more rapid peaks with greater overall periodicity and thus are more difficult to monitor (147, 148). However, there are no data demonstrating that increased periodicity is harmful otherwise.

Clinicians can use serum estradiol levels to monitor oral, transdermal, and intramuscular estradiol. Blood tests cannot monitor conjugated estrogens or synthetic estrogen use. Clinicians should measure serum estradiol and serum testosterone and maintain them at the level for premenopausal females (100 to 200 pg/mL and <50 ng/dL, respectively). The transdermal preparations and injectable estradiol cypionate or valerate preparations may confer an advantage in older transgender females who may be at higher risk for thromboembolic disease (149).

Values

Our recommendation to maintain levels of gender-affirming hormones in the normal adult range places a high value on the avoidance of the long-term complications of pharmacologic doses. Those patients receiving endocrine treatment who have relative contraindications to hormones should have an in-depth discussion with their physician to balance the risks and benefits of therapy.

Remarks

Clinicians should inform all endocrine-treated individuals of all risks and benefits of gender-affirming hormones prior to initiating therapy. Clinicians should strongly encourage tobacco use cessation in transgender females to avoid increased risk of VTE and cardiovascular complications. We strongly discourage the unsupervised use of hormone therapy (150).

Not all individuals with GD/gender incongruence seek treatment as described (*e.g.*, male-to-eunuchs and individuals seeking partial transition). Tailoring current protocols to the individual may be done within the context of accepted safety guidelines using a multidisciplinary approach including mental health. No evidence-based protocols are available for these groups (151). We need prospective studies to better understand treatment options for these persons.

3.4. We suggest that endocrinologists provide education to transgender individuals undergoing treatment about the onset and time course of physical changes induced by sex hormone treatment. (2 |⊕○○○)

Evidence

Transgender males

Physical changes that are expected to occur during the first 1 to 6 months of testosterone therapy include cessation of menses, increased sexual desire, increased facial and body hair, increased oiliness of skin, increased muscle, and redistribution of fat mass. Changes that occur within the first year of testosterone therapy include deepening of the voice (152, 153), clitoromegaly, and male pattern hair loss (in some cases) (114, 144, 154, 155) (Table 12).

Transgender females

Physical changes that may occur in transgender females in the first 3 to 12 months of estrogen and antiandrogen therapy include decreased sexual desire, decreased spontaneous erections, decreased facial and body hair (usually mild), decreased oiliness of skin, increased breast tissue growth, and redistribution of fat mass (114, 139, 149, 154, 155, 161) (Table 13). Breast development is generally maximal at 2 years after initiating hormones (114, 139, 149, 155). Over a long period of time, the prostate gland and testicles will undergo atrophy.

Although the time course of breast development in transgender females has been studied (150), precise information about other changes induced by sex hormones is lacking (141). There is a great deal of variability among individuals, as evidenced during pubertal development. We all know that a major concern for transgender females is breast development. If we work with estrogens, the result will be often not what the transgender female expects.

Alternatively, there are transgender females who report an anecdotal improved breast development, mood, or sexual desire with the use of progestogens. However, there have been no well-designed studies of the role of progestogens in feminizing hormone regimens, so the question is still open.

Our knowledge concerning the natural history and effects of different cross-sex hormone therapies on breast

Table 12. Masculinizing Effects in Transgender Males

Effect	Onset	Maximum
Skin oiliness/acne	1–6 mo	1–2 y
Facial/body hair growth	6–12 mo	4–5 y
Scalp hair loss	6–12 mo	—[a]
Increased muscle mass/strength	6–12 mo	2–5 y
Fat redistribution	1–6 mo	2–5 y
Cessation of menses	1–6 mo	—[b]
Clitoral enlargement	1–6 mo	1–2 y
Vaginal atrophy	1–6 mo	1–2 y
Deepening of voice	6–12 mo	1–2 y

Estimates represent clinical observations: Toorians *et al.* (149), Asscheman *et al.* (156), Gooren *et al.* (157), Wierckx *et al.* (158).

[a] Prevention and treatment as recommended for biological men.

[b] Menorrhagia requires diagnosis and treatment by a gynecologist.

doi: 10.1210/jc.2017-01658

Table 13. Feminizing Effects in Transgender Females

Effect	Onset	Maximum
Redistribution of body fat	3–6 mo	2–3 y
Decrease in muscle mass and strength	3–6 mo	1–2 y
Softening of skin/decreased oiliness	3–6 mo	Unknown
Decreased sexual desire	1–3 mo	3–6 mo
Decreased spontaneous erections	1–3 mo	3–6 mo
Male sexual dysfunction	Variable	Variable
Breast growth	3–6 mo	2–3 y
Decreased testicular volume	3–6 mo	2–3 y
Decreased sperm production	Unknown	>3 y
Decreased terminal hair growth	6–12 mo	>3 y[a]
Scalp hair	Variable	—[b]
Voice changes	None	—[c]

Estimates represent clinical observations: Toorians *et al.* (149), Asscheman *et al.* (156), Gooren *et al.* (157).

[a]Complete removal of male sexual hair requires electrolysis or laser treatment or both.

[b]Familial scalp hair loss may occur if estrogens are stopped.

[c]Treatment by speech pathologists for voice training is most effective.

development in transgender females is extremely sparse and based on the low quality of evidence. Current evidence does not indicate that progestogens enhance breast development in transgender females, nor does evidence prove the absence of such an effect. This prevents us from drawing any firm conclusion at this moment and demonstrates the need for further research to clarify these important clinical questions (162).

Values and preferences

Transgender persons have very high expectations regarding the physical changes of hormone treatment and are aware that body changes can be enhanced by surgical procedures (*e.g.*, breast, face, and body habitus). Clear expectations for the extent and timing of sex hormone–induced changes may prevent the potential harm and expense of unnecessary procedures.

4.0 Adverse Outcome Prevention and Long-Term Care

Hormone therapy for transgender males and females confers many of the same risks associated with sex hormone replacement therapy in nontransgender persons. The risks arise from and are worsened by inadvertent or intentional use of supraphysiologic doses of sex hormones, as well as use of inadequate doses of sex hormones to maintain normal physiology (131, 139).

4.1. We suggest regular clinical evaluation for physical changes and potential adverse changes in response to sex steroid hormones and laboratory monitoring of sex steroid hormone levels every 3 months during the first year of hormone therapy for transgender males and females and then once or twice yearly. (2 ⊕⊕○○)

Evidence

Pretreatment screening and appropriate regular medical monitoring are recommended for both transgender males and females during the endocrine transition and periodically thereafter (26, 155). Clinicians should monitor weight and blood pressure, conduct physical exams, and assess routine health questions, such as tobacco use, symptoms of depression, and risk of adverse events such as deep vein thrombosis/pulmonary embolism and other adverse effects of sex steroids.

Transgender males

Table 14 contains a standard monitoring plan for transgender males on testosterone therapy (154, 159). Key issues include maintaining testosterone levels in the physiologic normal male range and avoiding adverse events resulting from excess testosterone therapy, particularly erythrocytosis, sleep apnea, hypertension, excessive weight gain, salt retention, lipid changes, and excessive or cystic acne (135).

Because oral 17-alkylated testosterone is not recommended, serious hepatic toxicity is not anticipated with parenteral or transdermal testosterone use (163, 164). Past concerns regarding liver toxicity with testosterone have been alleviated with subsequent reports that indicate the risk of serious liver disease is minimal (144, 165, 166).

Transgender females

Table 15 contains a standard monitoring plan for transgender females on estrogens, gonadotropin suppression, or antiandrogens (160). Key issues include avoiding supraphysiologic doses or blood levels of estrogen that may lead to increased risk for thromboembolic disease, liver dysfunction, and hypertension. Clinicians should monitor serum estradiol levels using laboratories participating in external quality control, as measurements of estradiol in blood can be very challenging (167).

VTE may be a serious complication. A study reported a 20-fold increase in venous thromboembolic disease in a large cohort of Dutch transgender subjects (161). This increase may have been associated with the use of the synthetic estrogen, ethinyl estradiol (149). The incidence decreased when clinicians stopped administering ethinyl estradiol (161). Thus, the use of synthetic estrogens and conjugated estrogens is undesirable because of the inability to regulate doses by measuring serum levels and the risk of thromboembolic disease. In a German gender clinic, deep vein thrombosis occurred in 1 of 60 of transgender females treated with a GnRH antagonist and oral

Table 14. Monitoring of Transgender Persons on Gender-Affirming Hormone Therapy: Transgender Male

1. Evaluate patient every 3 mo in the first year and then one to two times per year to monitor for appropriate signs of virilization and for development of adverse reactions.
2. Measure serum testosterone every 3 mo until levels are in the normal physiologic male range:[a]
   a. For testosterone enanthate/cypionate injections, the testosterone level should be measured midway between injections. The target level is 400–700 ng/dL to 400 ng/dL. Alternatively, measure peak and trough levels to ensure levels remain in the normal male range.
   b. For parenteral testosterone undecanoate, testosterone should be measured just before the following injection. If the level is <400 ng/dL, adjust dosing interval.
   c. For transdermal testosterone, the testosterone level can be measured no sooner than after 1 wk of daily application (at least 2 h after application).
3. Measure hematocrit or hemoglobin at baseline and every 3 mo for the first year and then one to two times a year. Monitor weight, blood pressure, and lipids at regular intervals.
4. Screening for osteoporosis should be conducted in those who stop testosterone treatment, are not compliant with hormone therapy, or who develop risks for bone loss.
5. If cervical tissue is present, monitoring as recommended by the American College of Obstetricians and Gynecologists.
6. Ovariectomy can be considered after completion of hormone transition.
7. Conduct sub- and periareolar annual breast examinations if mastectomy performed. If mastectomy is not performed, then consider mammograms as recommended by the American Cancer Society.

[a] Adapted from Lapauw *et al.* (154) and Ott *et al.* (159).

estradiol (141). The patient who developed a deep vein thrombosis was found to have a homozygous C677 T mutation in the methylenetetrahydrofolate reductase gene. In an Austrian gender clinic, administering gender-affirming hormones to 162 transgender females and 89 transgender males was not associated with VTE, despite an 8.0% and 5.6% incidence of thrombophilia (159). A more recent multinational study reported only 10 cases of VTE from a cohort of 1073 subjects (168). Thrombophilia screening of transgender persons initiating hormone treatment should be restricted to those with a personal or family history of VTE (159). Monitoring D-dimer levels during treatment is not recommended (169).

4.2. We suggest periodically monitoring prolactin levels in transgender females treated with estrogens. (2 l⊕⊕◯◯)

Evidence

Estrogen therapy can increase the growth of pituitary lactrotroph cells. There have been several reports of prolactinomas occurring after long-term, high-dose estrogen therapy (170–173). Up to 20% of transgender females treated with estrogens may have elevations in prolactin levels associated with enlargement of the pituitary gland (156). In most cases, the serum prolactin levels will return to the normal range with a reduction or discontinuation of the estrogen therapy or discontinuation of cyproterone acetate (157, 174, 175).

The onset and time course of hyperprolactinemia during estrogen treatment are not known. Clinicians should measure prolactin levels at baseline and then at least annually during the transition period and every 2 years thereafter. Given that only a few case studies reported prolactinomas, and prolactinomas were not reported in large cohorts of estrogen-treated persons, the risk is likely to be very low. Because the major presenting findings of microprolactinomas (hypogonadism and sometimes gynecomastia) are not apparent in transgender females, clinicians may perform radiologic examinations of the pituitary in those patients whose prolactin levels persistently increase despite stable or reduced estrogen levels. Some transgender individuals receive psychotropic medications that can increase prolactin levels (174).

Table 15. Monitoring of Transgender Persons on Gender-Affirming Hormone Therapy: Transgender Female

1. Evaluate patient every 3 mo in the first year and then one to two times per year to monitor for appropriate signs of feminization and for development of adverse reactions.
2. Measure serum testosterone and estradiol every 3 mo.
   a. Serum testosterone levels should be <50 ng/dL.
   b. Serum estradiol should not exceed the peak physiologic range: 100–200 pg/mL.
3. For individuals on spironolactone, serum electrolytes, particularly potassium, should be monitored every 3 mo in the first year and annually thereafter.
4. Routine cancer screening is recommended, as in nontransgender individuals (all tissues present).
5. Consider BMD testing at baseline (160). In individuals at low risk, screening for osteoporosis should be conducted at age 60 years or in those who are not compliant with hormone therapy.

This table presents strong recommendations and does not include lower level recommendations.

doi: 10.1210/jc.2017-01658

4.3. We suggest that clinicians evaluate transgender persons treated with hormones for cardiovascular risk factors using fasting lipid profiles, diabetes screening, and/or other diagnostic tools. (2 |⊕⊕◯◯)

Evidence

Transgender males

Administering testosterone to transgender males results in a more atherogenic lipid profile with lowered high-density lipoprotein cholesterol and higher triglyceride and low-density lipoprotein cholesterol values (176–179). Studies of the effect of testosterone on insulin sensitivity have mixed results (178, 180). A randomized, open-label uncontrolled safety study of transgender males treated with testosterone undecanoate demonstrated no insulin resistance after 1 year (181, 182). Numerous studies have demonstrated the effects of sex hormone treatment on the cardiovascular system (160, 179, 183, 184). Long-term studies from The Netherlands found no increased risk for cardiovascular mortality (161). Likewise, a meta-analysis of 19 randomized trials in nontransgender males on testosterone replacement showed no increased incidence of cardiovascular events (185). A systematic review of the literature found that data were insufficient (due to very low–quality evidence) to allow a meaningful assessment of patient-important outcomes, such as death, stroke, myocardial infarction, or VTE in transgender males (176). Future research is needed to ascertain the potential harm of hormonal therapies (176). Clinicians should manage cardiovascular risk factors as they emerge according to established guidelines (186).

Transgender females

A prospective study of transgender females found favorable changes in lipid parameters with increased high-density lipoprotein and decreased low-density lipoprotein concentrations (178). However, increased weight, blood pressure, and markers of insulin resistance attenuated these favorable lipid changes. In a meta-analysis, only serum triglycerides were higher at ≥24 months without changes in other parameters (187). The largest cohort of transgender females (mean age 41 years, followed for a mean of 10 years) showed no increase in cardiovascular mortality despite a 32% rate of tobacco use (161).

Thus, there is limited evidence to determine whether estrogen is protective or detrimental on lipid and glucose metabolism in transgender females (176). With aging, there is usually an increase of body weight. Therefore, as with nontransgender individuals, clinicians should monitor and manage glucose and lipid metabolism and blood pressure regularly according to established guidelines (186).

4.4. We recommend that clinicians obtain BMD measurements when risk factors for osteoporosis exist, specifically in those who stop sex hormone therapy after gonadectomy. (1 |⊕⊕◯◯)

Evidence

Transgender males

Baseline bone mineral measurements in transgender males are generally in the expected range for their pretreatment gender (188). However, adequate dosing of testosterone is important to maintain bone mass in transgender males (189, 190). In one study (190), serum LH levels were inversely related to BMD, suggesting that low levels of sex hormones were associated with bone loss. Thus, LH levels in the normal range may serve as an indicator of the adequacy of sex steroid administration to preserve bone mass. The protective effect of testosterone may be mediated by peripheral conversion to estradiol, both systemically and locally in the bone.

Transgender females

A baseline study of BMD reported T scores less than −2.5 in 16% of transgender females (191). In aging males, studies suggest that serum estradiol more positively correlates with BMD than does testosterone (192, 193) and is more important for peak bone mass (194). Estrogen preserves BMD in transgender females who continue on estrogen and antiandrogen therapies (188, 190, 191, 195, 196).

Fracture data in transgender males and females are not available. Transgender persons who have undergone gonadectomy may choose not to continue consistent sex steroid treatment after hormonal and surgical sex reassignment, thereby becoming at risk for bone loss. There have been no studies to determine whether clinicians should use the sex assigned at birth or affirmed gender for assessing osteoporosis (*e.g.*, when using the FRAX tool). Although some researchers use the sex assigned at birth (with the assumption that bone mass has usually peaked for transgender people who initiate hormones in early adulthood), this should be assessed on a case-by-case basis until there are more data available. This assumption will be further complicated by the increasing prevalence of transgender people who undergo hormonal transition at a pubertal age or soon after puberty. Sex for comparison within risk assessment tools may be based on the age at which hormones were initiated and the length of exposure to hormones. In some cases, it may be

reasonable to assess risk using both the male and female calculators and using an intermediate value. Because all subjects underwent normal pubertal development, with known effects on bone size, reference values for birth sex were used for all participants (154).

4.5. We suggest that transgender females with no known increased risk of breast cancer follow breast-screening guidelines recommended for those designated female at birth. (2 |⊕⊕◯◯)

4.6. We suggest that transgender females treated with estrogens follow individualized screening according to personal risk for prostatic disease and prostate cancer. (2 |⊕◯◯◯)

Evidence

Studies have reported a few cases of breast cancer in transgender females (197–200). A Dutch study of 1800 transgender females followed for a mean of 15 years (range of 1 30 years) found one case of breast cancer. The Women's Health Initiative study reported that females taking conjugated equine estrogen without progesterone for 7 years did not have an increased risk of breast cancer as compared with females taking placebo (137).

In transgender males, a large retrospective study conducted at the U.S. Veterans Affairs medical health system identified seven breast cancers (194). The authors reported that this was not above the expected rate of breast cancers in cisgender females in this cohort. Furthermore, they did report one breast cancer that developed in a transgender male patient after mastectomy, supporting the fact that breast cancer can occur even after mastectomy. Indeed, there have been case reports of breast cancer developing in subareolar tissue in transgender males, which occurred after mastectomy (201, 202).

Women with primary hypogonadism (Turner syndrome) treated with estrogen replacement exhibited a significantly decreased incidence of breast cancer as compared with national standardized incidence ratios (203, 204). These studies suggest that estrogen therapy does not increase the risk of breast cancer in the short term (<20 to 30 years). We need long-term studies to determine the actual risk, as well as the role of screening mammograms. Regular examinations and gynecologic advice should determine monitoring for breast cancer.

Prostate cancer is very rare before the age of 40, especially with androgen deprivation therapy (205). Childhood or pubertal castration results in regression of the prostate and adult castration reverses benign prostate hypertrophy (206). Although van Kesteren *et al.* (207) reported that estrogen therapy does not induce hypertrophy or premalignant changes in the prostates of transgender females, studies have reported cases of benign prostatic hyperplasia in transgender females treated with estrogens for 20 to 25 years (208, 209). Studies have also reported a few cases of prostate carcinoma in transgender females (210–214).

Transgender females may feel uncomfortable scheduling regular prostate examinations. Gynecologists are not trained to screen for prostate cancer or to monitor prostate growth. Thus, it may be reasonable for transgender females who transitioned after age 20 years to have annual screening digital rectal examinations after age 50 years and prostate-specific antigen tests consistent with U.S. Preventive Services Task Force Guidelines (215).

4.7. We advise that clinicians determine the medical necessity of including a total hysterectomy and oophorectomy as part of gender-affirming surgery. (Ungraded Good Practice Statement)

Evidence

Although aromatization of testosterone to estradiol in transgender males has been suggested as a risk factor for endometrial cancer (216), no cases have been reported. When transgender males undergo hysterectomy, the uterus is small and there is endometrial atrophy (217, 218). Studies have reported cases of ovarian cancer (219, 220). Although there is limited evidence for increased risk of reproductive tract cancers in transgender males, health care providers should determine the medical necessity of a laparoscopic total hysterectomy as part of a gender-affirming surgery to prevent reproductive tract cancer (221).

Values

Given the discomfort that transgender males experience accessing gynecologic care, our recommendation for the medical necessity of total hysterectomy and oophorectomy places a high value on eliminating the risks of female reproductive tract disease and cancer and a lower value on avoiding the risks of these surgical procedures (related to the surgery and to the potential undesirable health consequences of oophorectomy) and their associated costs.

Remarks

The sexual orientation and type of sexual practices will determine the need and types of gynecologic care required following transition. Additionally, in certain countries, the approval required to change the sex in a birth certificate for transgender males may be dependent on having a complete hysterectomy. Clinicians should help patients research nonmedical administrative criteria and

doi: 10.1210/jc.2017-01658

provide counseling. If individuals decide not to undergo hysterectomy, screening for cervical cancer is the same as all other females.

5.0 Surgery for Sex Reassignment and Gender Confirmation

For many transgender adults, genital gender-affirming surgery may be the necessary step toward achieving their ultimate goal of living successfully in their desired gender role. The type of surgery falls into two main categories: (1) those that directly affect fertility and (2) those that do not. Those that change fertility (previously called sex reassignment surgery) include genital surgery to remove the penis and gonads in the male and removal of the uterus and gonads in the female. The surgeries that effect fertility are often governed by the legal system of the state or country in which they are performed. Other gender-conforming surgeries that do not directly affect fertility are not so tightly governed.

Gender-affirming surgical techniques have improved markedly during the past 10 years. Reconstructive genital surgery that preserves neurologic sensation is now the standard. The satisfaction rate with surgical reassignment of sex is now very high (187). Additionally, the mental health of the individual seems to be improved by participating in a treatment program that defines a pathway of gender-affirming treatment that includes hormones and surgery (130, 144) (Table 16).

Surgery that affects fertility is irreversible. The World Professional Association for Transgender Health Standards of Care (222) emphasizes that the "threshold of 18 should not be seen as an indication in itself for active intervention." If the social transition has not been satisfactory, if the person is not satisfied with or is ambivalent about the effects of sex hormone treatment, or if the person is ambivalent about surgery then the individual should not be referred for surgery (223, 224).

Gender-affirming genital surgeries for transgender females that affect fertility include gonadectomy, penectomy, and creation of a neovagina (225, 226). Surgeons often invert the skin of the penis to form the wall of the vagina, and several literatures reviews have reported on outcomes (227). Sometimes there is inadequate tissue to form a full neovagina, so clinicians have revisited using intestine and found it to be successful (87, 228, 229). Some newer vaginoplasty techniques may involve autologuous oral epithelial cells (230, 231).

The scrotum becomes the labia majora. Surgeons use reconstructive surgery to fashion the clitoris and its hood, preserving the neurovascular bundle at the tip of the penis as the neurosensory supply to the clitoris. Some surgeons are also creating a sensate pedicled-spot adding a G spot to the neovagina to increase sensation (232). Most recently, plastic surgeons have developed techniques to fashion labia minora. To further complete the feminization, uterine transplants have been proposed and even attempted (233).

Neovaginal prolapse, rectovaginal fistula, delayed healing, vaginal stenosis, and other complications do sometimes occur (234, 235). Clinicians should strongly remind the transgender person to use their dilators to maintain the depth and width of the vagina throughout the postoperative period. Genital sexual responsivity and other aspects of sexual function are usually preserved following genital gender-affirming surgery (236, 237).

Ancillary surgeries for more feminine or masculine appearance are not within the scope of this guideline. Voice therapy by a speech language pathologist is available to transform speech patterns to the affirmed gender (148). Spontaneous voice deepening occurs during testosterone treatment of transgender males (152, 238). No studies have compared the effectiveness of speech therapy, laryngeal surgery, or combined treatment.

Breast surgery is a good example of gender-confirming surgery that does not affect fertility. In all females, breast size exhibits a very broad spectrum. For transgender females to make the best informed decision, clinicians should delay breast augmentation surgery until the patient has completed at least 2 years of estrogen therapy, because the breasts continue to grow during that time (141, 155).

Another major procedure is the removal of facial and masculine-appearing body hair using either electrolysis or

Table 16. Criteria for Gender-Affirming Surgery, Which Affects Fertility

1. Persistent, well-documented gender dysphoria
2. Legal age of majority in the given country
3. Having continuously and responsibly used gender-affirming hormones for 12 mo (if there is no medical contraindication to receiving such therapy)
4. Successful continuous full-time living in the new gender role for 12 mo
5. If significant medical or mental health concerns are present, they must be well controlled
6. Demonstrable knowledge of all practical aspects of surgery (*e.g.*, cost, required lengths of hospitalizations, likely complications, postsurgical rehabilitation)

laser treatments. Other feminizing surgeries, such as that to feminize the face, are now becoming more popular (239–241).

In transgender males, clinicians usually delay gender-affirming genital surgeries until after a few years of androgen therapy. Those surgeries that affect fertility in this group include oophorectomy, vaginectomy, and complete hysterectomy. Surgeons can safely perform them vaginally with laparoscopy. These are sometimes done in conjunction with the creation of a neopenis. The cosmetic appearance of a neopenis is now very good, but the surgery is multistage and very expensive (242, 243). Radial forearm flap seems to be the most satisfactory procedure (228, 244). Other flaps also exist (245). Surgeons can make neopenile erections possible by reinervation of the flap and subsequent contraction of the muscle, leading to stiffening of the neopenis (246, 247), but results are inconsistent (248). Surgeons can also stiffen the penis by imbedding some mechanical device (*e.g.*, a rod or some inflatable apparatus) (249, 250). Because of these limitations, the creation of a neopenis has often been less than satisfactory. Recently, penis transplants are being proposed (233).

In fact, most transgender males do not have any external genital surgery because of the lack of access, high cost, and significant potential complications. Some choose a metaoidioplasty that brings forward the clitoris, thereby allowing them to void in a standing position without wetting themselves (251, 252). Surgeons can create the scrotum from the labia majora with good cosmetic effect and can implant testicular prostheses (253).

The most important masculinizing surgery for the transgender male is mastectomy, and it does not affect fertility. Breast size only partially regresses with androgen therapy (155). In adults, discussions about mastectomy usually take place after androgen therapy has started. Because some transgender male adolescents present after significant breast development has occurred, they may also consider mastectomy 2 years after they begin androgen therapy and before age 18 years. Clinicians should individualize treatment based on the physical and mental health status of the individual. There are now newer approaches to mastectomy with better outcomes (254, 255). These often involve chest contouring (256). Mastectomy is often necessary for living comfortably in the new gender (256).

5.1. We recommend that a patient pursue genital gender-affirming surgery only after the MHP and the clinician responsible for endocrine transition therapy both agree that surgery is medically necessary and would benefit the patient's overall health and/or well-being. (1 ⊕⊕◯◯)

5.2. We advise that clinicians approve genital gender-affirming surgery only after completion of at least 1 year of consistent and compliant hormone treatment, unless hormone therapy is not desired or medically contraindicated. (Ungraded Good Practice Statement)

5.3. We advise that the clinician responsible for endocrine treatment and the primary care provider ensure appropriate medical clearance of transgender individuals for genital gender-affirming surgery and collaborate with the surgeon regarding hormone use during and after surgery. (Ungraded Good Practice Statement)

5.4. We recommend that clinicians refer hormone-treated transgender individuals for genital surgery when: (1) the individual has had a satisfactory social role change, (2) the individual is satisfied about the hormonal effects, and (3) the individual desires definitive surgical changes. (1 ⊕◯◯◯)

5.5. We suggest that clinicians delay gender-affirming genital surgery involving gonadectomy and/or hysterectomy until the patient is at least 18 years old or legal age of majority in his or her country. (2 ⊕⊕◯◯).

5.6. We suggest that clinicians determine the timing of breast surgery for transgender males based upon the physical and mental health status of the individual. There is insufficient evidence to recommend a specific age requirement. (2 ⊕◯◯◯)

Evidence

Owing to the lack of controlled studies, incomplete follow-up, and lack of valid assessment measures, evaluating various surgical approaches and techniques is difficult. However, one systematic review including a large numbers of studies reported satisfactory cosmetic and functional results for vaginoplasty/neovagina construction (257). For transgender males, the outcomes are less certain. However, the problems are now better understood (258). Several postoperative studies report significant long-term psychological and psychiatric pathology (259–261). One study showed satisfaction with breasts, genitals, and femininity increased significantly and showed the importance of surgical treatment as a key therapeutic option for transgender females (262). Another analysis demonstrated that, despite the young average age at death following surgery and the relatively larger number of individuals with somatic morbidity, the study does not allow for determination of

doi: 10.1210/jc.2017-01658

causal relationships between, for example, specific types of hormonal or surgical treatment received and somatic morbidity and mortality (263). Reversal surgery in regretful male-to-female transsexuals after sexual reassignment surgery represents a complex, multistage procedure with satisfactory outcomes. Further insight into the characteristics of persons who regret their decision postoperatively would facilitate better future selection of applicants eligible for sexual reassignment surgery. We need more studies with appropriate controls that examine long-term quality of life, psychosocial outcomes, and psychiatric outcomes to determine the long-term benefits of surgical treatment.

When a transgender individual decides to have gender-affirming surgery, both the hormone prescribing clinician and the MHP must certify that the patient satisfies criteria for gender-affirming surgery (Table 16).

There is some concern that estrogen therapy may cause an increased risk for venous thrombosis during or following surgery (176). For this reason, the surgeon and the hormone-prescribing clinician should collaborate in making a decision about the use of hormones before and following surgery. One study suggests that preoperative factors (such as compliance) are less important for patient satisfaction than are the physical postoperative results (56). However, other studies and clinical experience dictate that individuals who do not follow medical instructions and do not work with their physicians toward a common goal do not achieve treatment goals (264) and experience higher rates of postoperative infections and other complications (265, 266). It is also important that the person requesting surgery feels comfortable with the anatomical changes that have occurred during hormone therapy. Dissatisfaction with social and physical outcomes during the hormone transition may be a contraindication to surgery (223).

An endocrinologist or experienced medical provider should monitor transgender individuals after surgery. Those who undergo gonadectomy will require hormone replacement therapy, surveillance, or both to prevent adverse effects of chronic hormone deficiency.

Financial Disclosures of the Task Force*

Wylie C. Hembree (chair)—financial or business/organizational interests: none declared, significant financial interest or leadership position: none declared. **Peggy T. Cohen-Kettenis**—financial or business/organizational interests: none declared, significant financial interest or leadership position: none declared. **Louis Gooren**—financial or business/organizational interests: none declared, significant financial interest or leadership position: none declared. **Sabine E. Hannema**—financial or business/organizational interests: none declared, significant financial interest or leadership position: Ferring Pharmaceuticals Inc. (lecture/conference), Pfizer (lecture). **Walter J. Meyer**—financial or business/organizational interests: none declared, significant financial interest or leadership position: none declared. **M. Hassan Murad****—financial or business/organizational interests: Mayo Clinic, Evidence-based Practice Center, significant financial interest or leadership position: none declared. **Stephen M. Rosenthal**—financial or business/organizational interests: AbbVie (consultant), National Institutes of Health (grantee), significant financial interest or leadership position: Pediatric Endocrine Society (immediate past president). **Joshua D. Safer, FACP**—financial or business/organizational interests: none declared, significant financial interest or leadership position: none declared. **Vin Tangpricha**—financial or business/organizational interests: Cystic Fibrosis Foundation (grantee), National Institutes of Health (grantee), significant financial interest or leadership position, Elsevier *Journal of Clinical and Translational Endocrinology* (editor). **Guy G. T'Sjoen**—financial or business/organizational interests: none declared, significant financial interest or leadership position: none declared.* Financial, business, and organizational disclosures of the task force cover the year prior to publication. Disclosures prior to this time period are archived.**Evidence-based reviews for this guideline were prepared under contract with the Endocrine Society.

Acknowledgments

Correspondence and Reprint Requests: The Endocrine Society, 2055 L Street NW, Suite 600, Washington, DC 20036. E-mail: publications@endocrine.org; Phone: 202971-3636.

Disclosure Summary: See Financial Disclosures.

Disclaimer: The Endocrine Society's clinical practice guidelines are developed to be of assistance to endocrinologists by providing guidance and recommendations for particular areas of practice. The guidelines should not be considered inclusive of all proper approaches or methods, or exclusive of others. The guidelines cannot guarantee any specific outcome, nor do they establish a standard of care. The guidelines are not intended to dictate the treatment of a particular patient. Treatment decisions must be made based on the independent judgement of healthcare providers and each patient's individual circumstances.

The Endocrine Society makes no warranty, express or implied, regarding the guidelines and specifically excludes any warranties of merchantability and fitness for a particular use or purpose. The Society shall not be liable for direct, indirect,

special, incidental, or consequential damages related to the use of the information contained herein.

References

1. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, Guyatt GH, Harbour RT, Haugh MC, Henry D, Hill S, Jaeschke R, Leng G, Liberati A, Magrini N, Mason J, Middleton P, Mrukowicz J, O'Connell D, Oxman AD, Phillips B, Schünemann HJ, Edejer T, Varonen H, Vist GE, Williams JW, Jr, Zaza S; GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ*. 2004;**328**(7454):1490.
2. Swiglo BA, Murad MH, Schünemann HJ, Kunz R, Vigersky RA, Guyatt GH, Montori VM. A case for clarity, consistency, and helpfulness: state-of-the-art clinical practice guidelines in endocrinology using the grading of recommendations, assessment, development, and evaluation system. *J Clin Endocrinol Metab*. 2008;**93**(3):666–673.
3. Bullough VL. Transsexualism in history. *Arch Sex Behav*. 1975; **4**(5):561–571.
4. Benjamin H. The transsexual phenomenon. *Trans N Y Acad Sci*. 1967;**29**(4):428–430.
5. Meyerowitz J. *How Sex Changed: A History of Transsexuality in the United States*. Cambridge, MA: Harvard University Press; 2002.
6. Hirschfeld M. *Was muss das Volk vom Dritten Geschlecht wissen*. Verlag Max Spohr, Leipzig; 1901.
7. Fisk NM. Editorial: Gender dysphoria syndrome—the conceptualization that liberalizes indications for total gender reorientation and implies a broadly based multi-dimensional rehabilitative regimen. *West J Med*. 1974;**120**(5):386–391.
8. Diamond L. Transgender experience and identity. In: Schwartz SJ, Luyckx K, Vignoles VL, eds. *Handbook of Identity Theory and Research*. New York, NY: Springer; 2011:629–647.
9. Queen C, Schimel L, eds. *PoMoSexuals: Challenging Assumptions About Gender and Sexuality*. San Francisco, CA: Cleis Press; 1997.
10. Case LK, Ramachandran VS. Alternating gender incongruity: a new neuropsychiatric syndrome providing insight into the dynamic plasticity of brain-sex. *Med Hypotheses*. 2012;**78**(5): 626–631.
11. Johnson TW, Wassersug RJ. Gender identity disorder outside the binary: when gender identity disorder-not otherwise specified is not good enough. *Arch Sex Behav*. 2010;**39**(3):597–598.
12. Wibowo E, Wassersug R, Warkentin K, Walker L, Robinson J, Brotto L, Johnson T. Impact of androgen deprivation therapy on sexual function: a response. *Asian J Androl*. 2012;**14**(5):793–794.
13. Pasquesoone V. 7 countries giving transgender people fundamental rights the U.S. still won't. 2014. Available at: https://mic.com/articles/87149/7-countries-giving-transgender-people-fundamental-rights-the-u-s-still-won-t. Accessed 26 August 2016.
14. American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*. 5th ed. Arlington, VA: American Psychiatric Association Publishing.
15. Drescher J, Cohen-Kettenis P, Winter S. Minding the body: situating gender identity diagnoses in the ICD-11. *Int Rev Psychiatry*. 2012;**24**(6):568–577.
16. World Professional Association for Transgender Health. Standards of care for the health of transsexual, transgender, and gender nonconforming people. Available at: http://www.wpath.org/site_page.cfm?pk_association_webpage_menu=1351&pk_association_webpage=3926. Accessed 1 September 2017.
17. Kreukels BP, Haraldsen IR, De Cuypere G, Richter-Appelt H, Gijs L, Cohen-Kettenis PT. A European network for the investigation of gender incongruence: the ENIGI initiative. *Eur Psychiatry*. 2012;**27**(6):445–450.
18. Dekker MJ, Wierckx K, Van Caenegem E, Klaver M, Kreukels BP, Elaut E, Fisher AD, van Trotsenburg MA, Schreiner T, den Heijer M, T'Sjoen G. A European network for the investigation of gender incongruence: endocrine part. *J Sex Med*. 2016;**13**(6):994–999.
19. Ruble DN, Martin CL, Berenbaum SA. Gender development. In: Damon WL, Lerner RM, Eisenberg N, eds. *Handbook of Child Psychology: Social, Emotional, and Personality Development*. **Vol. 3**. 6th ed. New York, NY: Wiley; 2006;858–931.
20. Steensma TD, Kreukels BP, de Vries AL, Cohen-Kettenis PT. Gender identity development in adolescence. *Horm Behav*. 2013; **64**(2):288–297.
21. Rosenthal SM. Approach to the patient: transgender youth: endocrine considerations. *J Clin Endocrinol Metab*. 2014;**99**(12): 4379–4389.
22. Saraswat A, Weinand JD, Safer JD. Evidence supporting the biologic nature of gender identity. *Endocr Pract*. 2015;**21**(2): 199–204.
23. Gooren L. The biology of human psychosexual differentiation. *Horm Behav*. 2006;**50**(4):589–601.
24. Berenbaum SA, Meyer-Bahlburg HF. Gender development and sexuality in disorders of sex development. *Horm Metab Res*. 2015; **47**(5):361–366.
25. Dessens AB, Slijper FME, Drop SLS. Gender dysphoria and gender change in chromosomal females with congenital adrenal hyperplasia. *Arch Sex Behav*. 2005;**34**(4):389–397.
26. Meyer-Bahlburg HFL, Dolezal C, Baker SW, Ehrhardt AA, New MI. Gender development in women with congenital adrenal hyperplasia as a function of disorder severity. *Arch Sex Behav*. 2006; **35**(6):667–684.
27. Frisén L, Nordenström A, Falhammar H, Filipsson H, Holmdahl G, Janson PO, Thorén M, Hagenfeldt K, Möller A, Nordenskjöld A. Gender role behavior, sexuality, and psychosocial adaptation in women with congenital adrenal hyperplasia due to CYP21A2 deficiency. *J Clin Endocrinol Metab*. 2009;**94**(9):3432–3439.
28. Meyer-Bahlburg HFL, Dolezal C, Baker SW, Carlson AD, Obeid JS, New MI. Prenatal androgenization affects gender-related behavior but not gender identity in 5–12-year-old girls with congenital adrenal hyperplasia. *Arch Sex Behav*. 2004;**33**(2):97–104.
29. Cohen-Kettenis PT. Gender change in 46,XY persons with 5α-reductase-2 deficiency and 17β-hydroxysteroid dehydrogenase-3 deficiency. *Arch Sex Behav*. 2005;**34**(4):399–410.
30. Reiner WG, Gearhart JP. Discordant sexual identity in some genetic males with cloacal exstrophy assigned to female sex at birth. *N Engl J Med*. 2004;**350**(4):333–341.
31. Meyer-Bahlburg HFL. Gender identity outcome in female-raised 46,XY persons with penile agenesis, cloacal exstrophy of the bladder, or penile ablation. *Arch Sex Behav*. 2005;**34**(4):423–438.
32. Coolidge FL, Thede LL, Young SE. The heritability of gender identity disorder in a child and adolescent twin sample. *Behav Genet*. 2002;**32**(4):251–257.
33. Heylens G, De Cuypere G, Zucker KJ, Schelfaut C, Elaut E, Vanden Bossche H, De Baere E, T'Sjoen G. Gender identity disorder in twins: a review of the case report literature. *J Sex Med*. 2012;**9**(3):751–757.
34. Fernández R, Esteva I, Gómez-Gil E, Rumbo T, Almaraz MC, Roda E, Haro-Mora J-J, Guillamón A, Pásaro E. Association study of ERβ, AR, and CYP19A1 genes and MtF transsexualism. *J Sex Med*. 2014;**11**(12):2986–2994.
35. Henningsson S, Westberg L, Nilsson S, Lundström B, Ekselius L, Bodlund O, Lindström E, Hellstrand M, Rosmond R, Eriksson E, Landén M. Sex steroid-related genes and male-to-female transsexualism. *Psychoneuroendocrinology*. 2005;**30**(7):657–664.
36. Hare L, Bernard P, Sánchez FJ, Baird PN, Vilain E, Kennedy T, Harley VR. Androgen receptor repeat length polymorphism associated with male-to-female transsexualism. *Biol Psychiatry*. 2009;**65**(1):93–96.
37. Lombardo F, Toselli L, Grassetti D, Paoli D, Masciandaro P, Valentini F, Lenzi A, Gandini L. Hormone and genetic study in

male to female transsexual patients. *J Endocrinol Invest*. 2013; **36**(8):550–557.
38. Ujike H, Otani K, Nakatsuka M, Ishii K, Sasaki A, Oishi T, Sato T, Okahisa Y, Matsumoto Y, Namba Y, Kimata Y, Kuroda S. Association study of gender identity disorder and sex hormone-related genes. *Prog Neuropsychopharmacol Biol Psychiatry*. 2009;**33**(7):1241–1244.
39. Kreukels BP, Guillamon A. Neuroimaging studies in people with gender incongruence. *Int Rev Psychiatry*. 2016;**28**(1): 120–128.
40. Steensma TD, Biemond R, de Boer F, Cohen-Kettenis PT. Desisting and persisting gender dysphoria after childhood: a qualitative follow-up study. *Clin Child Psychol Psychiatry*. 2011;**16**(4):499–516.
41. Wallien MSC, Cohen-Kettenis PT. Psychosexual outcome of gender-dysphoric children. *J Am Acad Child Adolesc Psychiatry*. 2008;**47**(12):1413–1423.
42. Steensma TD, McGuire JK, Kreukels BPC, Beekman AJ, Cohen-Kettenis PT. Factors associated with desistence and persistence of childhood gender dysphoria: a quantitative follow-up study. *J Am Acad Child Adolesc Psychiatry*. 2013;**52**(6):582–590.
43. Cohen-Kettenis PT, Owen A, Kaijser VG, Bradley SJ, Zucker KJ. Demographic characteristics, social competence, and behavior problems in children with gender identity disorder: a cross-national, cross-clinic comparative analysis. *J Abnorm Child Psychol*. 2003;**31**(1):41–53.
44. Dhejne C, Van Vlerken R, Heylens G, Arcelus J. Mental health and gender dysphoria: a review of the literature. *Int Rev Psychiatry*. 2016;**28**(1):44–57.
45. Pasterski V, Gilligan L, Curtis R. Traits of autism spectrum disorders in adults with gender dysphoria. *Arch Sex Behav*. 2014; **43**(2):387–393.
46. Spack NP, Edwards-Leeper L, Feldman HA, Leibowitz S, Mandel F, Diamond DA, Vance SR. Children and adolescents with gender identity disorder referred to a pediatric medical center. *Pediatrics*. 2012;**129**(3):418–425.
47. Terada S, Matsumoto Y, Sato T, Okabe N, Kishimoto Y, Uchitomi Y. Factors predicting psychiatric co-morbidity in gender-dysphoric adults. *Psychiatry Res*. 2012;**200**(2-3):469–474.
48. VanderLaan DP, Leef JH, Wood H, Hughes SK, Zucker KJ. Autism spectrum disorder risk factors and autistic traits in gender dysphoric children. *J Autism Dev Disord*. 2015;**45**(6):1742–1750.
49. de Vries ALC, Doreleijers TAH, Steensma TD, Cohen-Kettenis PT. Psychiatric comorbidity in gender dysphoric adolescents. *J Child Psychol Psychiatry*. 2011;**52**(11):1195–1202.
50. de Vries ALC, Noens ILJ, Cohen-Kettenis PT, van Berckelaer-Onnes IA, Doreleijers TA. Autism spectrum disorders in gender dysphoric children and adolescents. *J Autism Dev Disord*. 2010; **40**(8):930–936.
51. Wallien MSC, Swaab H, Cohen-Kettenis PT. Psychiatric comorbidity among children with gender identity disorder. *J Am Acad Child Adolesc Psychiatry*. 2007;**46**(10):1307–1314.
52. Kuiper AJ, Cohen-Kettenis PT. Gender role reversal among postoperative transsexuals. Available at: https://www.atria.nl/ezines/web/IJT/97-03/numbers/symposion/ijtc0502.htm. Accessed 26 August 2016.
53. Landén M, Wålinder J, Hambert G, Lundström B. Factors predictive of regret in sex reassignment. *Acta Psychiatr Scand*. 1998; **97**(4):284–289.
54. Olsson S-E, Möller A. Regret after sex reassignment surgery in a male-to-female transsexual: a long-term follow-up. *Arch Sex Behav*. 2006;**35**(4):501–506.
55. Pfäfflin F, Junge A, eds. *Geschlechtsumwandlung: Abhandlungen zur Transsexualität*. Stuttgart, Germany: Schattauer; 1992.
56. Lawrence AA. Factors associated with satisfaction or regret following male-to-female sex reassignment surgery. *Arch Sex Behav*. 2003;**32**(4):299–315.
57. Cohen-Kettenis PT, Pfäfflin F. *Transgenderism and Intersexuality in Childhood and Adolescence: Making Choices*. Thousand Oaks, CA: SAGE Publications; 2003.
58. Di Ceglie D, Freedman D, McPherson S, Richardson P. Children and adolescents referred to a specialist gender identity development service: clinical features and demographic characteristics. Available at: https://www.researchgate.net/publication/276061306_Children_and_Adolescents_Referred_to_a_Specialist_Gender_Identity_Development_Service_Clinical_Features_and_Demographic_Characteristics. Accessed 20 July 2017.
59. Gijs L, Brewaeys A. Surgical treatment of gender dysphoria in adults and adolescents: recent developments, effectiveness, and challenges. *Annu Rev Sex Res*. 2007;**18**:178–224.
60. Cohen-Kettenis PT, van Goozen SHM. Sex reassignment of adolescent transsexuals: a follow-up study. *J Am Acad Child Adolesc Psychiatry*. 1997;**36**(2):263–271.
61. Smith YLS, van Goozen SHM, Cohen-Kettenis PT. Adolescents with gender identity disorder who were accepted or rejected for sex reassignment surgery: a prospective follow-up study. *J Am Acad Child Adolesc Psychiatry*. 2001;**40**(4):472–481.
62. Smith YLS, Van Goozen SHM, Kuiper AJ, Cohen-Kettenis PT. Sex reassignment: outcomes and predictors of treatment for adolescent and adult transsexuals. *Psychol Med*. 2005;**35**(1):89–99.
63. de Vries ALC, McGuire JK, Steensma TD, Wagenaar ECF, Doreleijers TAH, Cohen-Kettenis PT. Young adult psychological outcome after puberty suppression and gender reassignment. *Pediatrics*. 2014;**134**(4):696–704.
64. Cole CM, O'Boyle M, Emory LE, Meyer WJ III. Comorbidity of gender dysphoria and other major psychiatric diagnoses. *Arch Sex Behav*. 1997;**26**(1):13–26.
65. Cohen-Kettenis PT, Schagen SEE, Steensma TD, de Vries ALC, Delemarre-van de Waal HA. Puberty suppression in a gender-dysphoric adolescent: a 22-year follow-up. *Arch Sex Behav*. 2011; **40**(4):843–847.
66. First MB. Desire for amputation of a limb: paraphilia, psychosis, or a new type of identity disorder. *Psychol Med*. 2005;**35**(6): 919–928.
67. Wierckx K, Van Caenegem E, Pennings G, Elaut E, Dedecker D, Van de Peer F, Weyers S, De Sutter P, T'Sjoen G. Reproductive wish in transsexual men. *Hum Reprod*. 2012;**27**(2):483–487.
68. Wierckx K, Stuyver I, Weyers S, Hamada A, Agarwal A, De Sutter P, T'Sjoen G. Sperm freezing in transsexual women. *Arch Sex Behav*. 2012;**41**(5):1069–1071.
69. Bertelloni S, Baroncelli GI, Ferdeghini M, Menchini-Fabris F, Saggese G. Final height, gonadal function and bone mineral density of adolescent males with central precocious puberty after therapy with gonadotropin-releasing hormone analogues. *Eur J Pediatr*. 2000;**159**(5):369–374.
70. Büchter D, Behre HM, Kliesch S, Nieschlag E. Pulsatile GnRH or human chorionic gonadotropin/human menopausal gonadotropin as effective treatment for men with hypogonadotropic hypogonadism: a review of 42 cases. *Eur J Endocrinol*. 1998; **139**(3):298–303.
71. Liu PY, Turner L, Rushford D, McDonald J, Baker HW, Conway AJ, Handelsman DJ. Efficacy and safety of recombinant human follicle stimulating hormone (Gonal-F) with urinary human chorionic gonadotrophin for induction of spermatogenesis and fertility in gonadotrophin-deficient men. *Hum Reprod*. 1999; **14**(6):1540–1545.
72. Pasquino AM, Pucarelli I, Accardo F, Demiraj V, Segni M, Di Nardo R. Long-term observation of 87 girls with idiopathic central precocious puberty treated with gonadotropin-releasing hormone analogs: impact on adult height, body mass index, bone mineral content, and reproductive function. *J Clin Endocrinol Metab*. 2008;**93**(1):190–195.
73. Magiakou MA, Manousaki D, Papadaki M, Hadjidakis D, Levidou G, Vakaki M, Papaefstathiou A, Lalioti N, Kanaka-Gantenbein C, Piaditis G, Chrousos GP, Dacou-Voutetakis C. The

efficacy and safety of gonadotropin-releasing hormone analog treatment in childhood and adolescence: a single center, long-term follow-up study. *J Clin Endocrinol Metab*. 2010;**95**(1):109–117.
74. Baba T, Endo T, Honnma H, Kitajima Y, Hayashi T, Ikeda H, Masumori N, Kamiya H, Moriwaka O, Saito T. Association between polycystic ovary syndrome and female-to-male transsexuality. *Hum Reprod*. 2007;**22**(4):1011–1016.
75. Spinder T, Spijkstra JJ, van den Tweel JG, Burger CW, van Kessel H, Hompes PGA, Gooren LJG. The effects of long term testosterone administration on pulsatile luteinizing hormone secretion and on ovarian histology in eugonadal female to male transsexual subjects. *J Clin Endocrinol Metab*. 1989;**69**(1):151–157.
76. Baba T, Endo T, Ikeda K, Shimizu A, Honnma H, Ikeda H, Masumori N, Ohmura T, Kiya T, Fujimoto T, Koizumi M, Saito T. Distinctive features of female-to-male transsexualism and prevalence of gender identity disorder in Japan. *J Sex Med*. 2011;**8**(6):1686–1693.
77. Vujovic S, Popovic S, Sbutega-Milosevic G, Djordjevic M, Gooren L. Transsexualism in Serbia: a twenty-year follow-up study. *J Sex Med*. 2009;**6**(4):1018–1023.
78. Ikeda K, Baba T, Noguchi H, Nagasawa K, Endo T, Kiya T, Saito T. Excessive androgen exposure in female-to-male transsexual persons of reproductive age induces hyperplasia of the ovarian cortex and stroma but not polycystic ovary morphology. *Hum Reprod*. 2013;**28**(2):453–461.
79. Trebay G. He's pregnant. You're speechles. New York Times. 22 June 2008.
80. Light AD, Obedin-Maliver J, Sevelius JM, Kerns JL. Transgender men who experienced pregnancy after female-to-male gender transitioning. *Obstet Gynecol*. 2014;**124**(6):1120–1127.
81. De Sutter P. Donor inseminations in partners of female-to-male transsexuals: should the question be asked? *Reprod Biomed Online*. 2003;**6**(3):382, author reply 282–283.
82. De Roo C, Tilleman K, T'Sjoen G, De Sutter P. Fertility options in transgender people. *Int Rev Psychiatry*. 2016;**28**(1):112–119.
83. Wennink JMB, Delemarre-van de Waal HA, Schoemaker R, Schoemaker H, Schoemaker J. Luteinizing hormone and follicle stimulating hormone secretion patterns in boys throughout puberty measured using highly sensitive immunoradiometric assays. *Clin Endocrinol (Oxf)*. 1989;**31**(5):551–564.
84. Cohen-Kettenis PT, Delemarre-van de Waal HA, Gooren LJG. The treatment of adolescent transsexuals: changing insights. *J Sex Med*. 2008;**5**(8):1892–1897.
85. Delemarre-van de Waal HA, Cohen-Kettenis PT. Clinical management of gender identity disorder in adolescents: a protocol on psychological and paediatric endocrinology aspects. *Eur J Endocrinol*. 2006;**155**:S131–S137.
86. de Vries ALC, Steensma TD, Doreleijers TAH, Cohen-Kettenis PT. Puberty suppression in adolescents with gender identity disorder: a prospective follow-up study. *J Sex Med*. 2011;**8**(8):2276–2283.
87. Bouman MB, van Zeijl MCT, Buncamper ME, Meijerink WJHJ, van Bodegraven AA, Mullender MG. Intestinal vaginoplasty revisited: a review of surgical techniques, complications, and sexual function. *J Sex Med*. 2014;**11**(7):1835–1847.
88. Carel JC, Eugster EA, Rogol A, Ghizzoni L, Palmert MR, Antoniazzi F, Berenbaum S, Bourguignon JP, Chrousos GP, Coste J, Deal S, de Vries L, Foster C, Heger S, Holland J, Jahnukainen K, Juul A, Kaplowitz P, Lahlou N, Lee MM, Lee P, Merke DP, Neely EK, Oostdijk W, Phillip M, Rosenfield RL, Shulman D, Styne D, Tauber M, Wit JM; ESPE-LWPES GnRH Analogs Consensus Conference Group. Consensus statement on the use of gonadotropin-releasing hormone analogs in children. *Pediatrics*. 2009;**123**(4):e752–e762.
89. Roth CL, Brendel L, Rückert C, Hartmann K. Antagonistic and agonistic GnRH analogue treatment of precocious puberty: tracking gonadotropin concentrations in urine. *Horm Res*. 2005;**63**(5):257–262.
90. Roth C. Therapeutic potential of GnRH antagonists in the treatment of precocious puberty. *Expert Opin Investig Drugs*. 2002;**11**(9):1253–1259.
91. Tuvemo T. Treatment of central precocious puberty. *Expert Opin Investig Drugs*. 2006;**15**(5):495–505.
92. Schagen SE, Cohen-Kettenis PT, Delemarre-van de Waal HA, Hannema SE. Efficacy and safety of gonadotropin-releasing hormone agonist treatment to suppress puberty in gender dysphoric adolescents. *J Sex Med*. 2016;**13**(7):1125–1132.
93. Manasco PK, Pescovitz OH, Feuillan PP, Hench KD, Barnes KM, Jones J, Hill SC, Loriaux DL, Cutler GB, Jr. Resumption of puberty after long term luteinizing hormone-releasing hormone agonist treatment of central precocious puberty. *J Clin Endocrinol Metab*. 1988;**67**(2):368–372.
94. Klink D, Caris M, Heijboer A, van Trotsenburg M, Rotteveel J. Bone mass in young adulthood following gonadotropin-releasing hormone analog treatment and cross-sex hormone treatment in adolescents with gender dysphoria. *J Clin Endocrinol Metab*. 2015;**100**(2):E270–E275.
95. Finkelstein JS, Klibanski A, Neer RM. A longitudinal evaluation of bone mineral density in adult men with histories of delayed puberty. *J Clin Endocrinol Metab*. 1996;**81**(3):1152–1155.
96. Bertelloni S, Baroncelli GI, Ferdeghini M, Perri G, Saggese G. Normal volumetric bone mineral density and bone turnover in young men with histories of constitutional delay of puberty. *J Clin Endocrinol Metab*. 1998;**83**(12):4280–4283.
97. Darelid A, Ohlsson C, Nilsson M, Kindblom JM, Mellström D, Lorentzon M. Catch up in bone acquisition in young adult men with late normal puberty. *J Bone Miner Res*. 2012;**27**(10):2198–2207.
98. Mittan D, Lee S, Miller E, Perez RC, Basler JW, Bruder JM. Bone loss following hypogonadism in men with prostate cancer treated with GnRH analogs. *J Clin Endocrinol Metab*. 2002;**87**(8):3656–3661.
99. Saggese G, Bertelloni S, Baroncelli GI, Battini R, Franchi G. Reduction of bone density: an effect of gonadotropin releasing hormone analogue treatment in central precocious puberty. *Eur J Pediatr*. 1993;**152**(9):717–720.
100. Neely EK, Bachrach LK, Hintz RL, Habiby RL, Slemenda CW, Feezle L, Pescovitz OH. Bone mineral density during treatment of central precocious puberty. *J Pediatr*. 1995;**127**(5):819–822.
101. Bertelloni S, Baroncelli GI, Sorrentino MC, Perri G, Saggese G. Effect of central precocious puberty and gonadotropin-releasing hormone analogue treatment on peak bone mass and final height in females. *Eur J Pediatr*. 1998;**157**(5):363–367.
102. Thornton P, Silverman LA, Geffner ME, Neely EK, Gould E, Danoff TM. Review of outcomes after cessation of gonadotropin-releasing hormone agonist treatment of girls with precocious puberty. *Pediatr Endocrinol Rev*. 2014;**11**(3):306–317.
103. Lem AJ, van der Kaay DC, Hokken-Koelega AC. Bone mineral density and body composition in short children born SGA during growth hormone and gonadotropin releasing hormone analog treatment. *J Clin Endocrinol Metab*. 2013;**98**(1):77–86.
104. Antoniazzi F, Zamboni G, Bertoldo F, Lauriola S, Mengarda F, Pietrobelli A, Tatò L. Bone mass at final height in precocious puberty after gonadotropin-releasing hormone agonist with and without calcium supplementation. *J Clin Endocrinol Metab*. 2003;**88**(3):1096–1101.
105. Calcaterra V, Mannarino S, Corana G, Codazzi AC, Mazzola A, Brambilla P, Larizza D. Hypertension during therapy with triptorelin in a girl with precocious puberty. *Indian J Pediatr*. 2013;**80**(10):884–885.
106. Siomou E, Kosmeri C, Pavlou M, Vlahos AP, Argyropoulou MI, Siamopoulou A. Arterial hypertension during treatment with triptorelin in a child with Williams-Beuren syndrome. *Pediatr Nephrol*. 2014;**29**(9):1633–1636.
107. Staphorsius AS, Kreukels BPC, Cohen-Kettenis PT, Veltman DJ, Burke SM, Schagen SEE, Wouters FM, Delemarre-van de Waal

HA, Bakker J. Puberty suppression and executive functioning: an fMRI-study in adolescents with gender dysphoria. *Psychoneuroendocrinology*. 2015;**56**:190–199.
108. Hough D, Bellingham M, Haraldsen IR, McLaughlin M, Rennie M, Robinson JE, Solbakk AK, Evans NP. Spatial memory is impaired by peripubertal GnRH agonist treatment and testosterone replacement in sheep. *Psychoneuroendocrinology*. 2017; **75**:173–182.
109. Collipp PJ, Kaplan SA, Boyle DC, Plachte F, Kogut MD. Constitutional Isosexual Precocious Puberty. *Am J Dis Child*. 1964; **108**:399–405.
110. Hahn HB, Jr, Hayles AB, Albert A. Medroxyprogesterone and constitutional precocious puberty. *Mayo Clin Proc*. 1964;**39**: 182–190.
111. Kaplan SA, Ling SM, Irani NG. Idiopathic isosexual precocity. *Am J Dis Child*. 1968;**116**(6):591–598.
112. Schoen EJ. Treatment of idiopathic precocious puberty in boys. *J Clin Endocrinol Metab*. 1966;**26**(4):363–370.
113. Gooren L. Hormone treatment of the adult transsexual patient. *Horm Res*. 2005;**64**(Suppl 2):31–36.
114. Moore E, Wisniewski A, Dobs A. Endocrine treatment of transsexual people: a review of treatment regimens, outcomes, and adverse effects. *J Clin Endocrinol Metab*. 2003;**88**(8):3467–3473.
115. Krueger RB, Hembree W, Hill M. Prescription of medroxyprogesterone acetate to a patient with pedophilia, resulting in Cushing's syndrome and adrenal insufficiency. *Sex Abuse*. 2006; **18**(2):227–228.
116. Lynch MM, Khandheria MM, Meyer WJ. Retrospective study of the management of childhood and adolescent gender identity disorder using medroxyprogesterone acetate. *Int J Transgenderism*. 2015;**16**:201–208.
117. Tack LJW, Craen M, Dhondt K, Vanden Bossche H, Laridaen J, Cools M. Consecutive lynestrenol and cross-sex hormone treatment in biological female adolescents with gender dysphoria: a retrospective analysis. *Biol Sex Differ*. 2016;**7**:14.
118. Hembree WC, Cohen-Kettenis P, Delemarre-van de Waal HA, Gooren LJ, Meyer WJ 3rd, Spack NP, Tangpricha V, Montori VM; Endocrine Society. Endocrine treatment of transsexual persons: an Endocrine Society clinical practice guideline. *J Clin Endocrinol Metab*. 2009;**94**(9):3132–3154.
119. Mann L, Harmoni R, Power C. Adolescent decision-making: the development of competence. *J Adolesc*. 1989;**12**(3):265–278.
120. Stultiëns L, Goffin T, Borry P, Dierickx K, Nys H. Minors and informed consent: a comparative approach. *Eur J Health Law*. 2007;**14**(1):21–46.
121. Arshagouni P. "But I'm an adult now … sort of". Adolescent consent in health care decision-making and the adolescent brain. Available at: http://digitalcommons.law.umaryland.edu/cgi/viewcontent.cgi?article=1124&context=jhclp. Accessed 25 June 2017.
122. NHS. Prescribing of cross-sex hormones as part of the gender identity development service for children and adolescents. Available at: https://www.england.nhs.uk/commissioning/wp-content/uploads/sites/12/2016/08/clinical-com-pol-16046p.pdf. Accessed 14 June 2017.
123. Ankarberg-Lindgren C, Kriström B, Norjavaara E. Physiological estrogen replacement therapy for puberty induction in girls: a clinical observational study. *Horm Res Paediatr*. 2014;**81**(4): 239–244.
124. Olson J, Schrager SM, Clark LF, Dunlap SL, Belzer M. Subcutaneous testosterone: an effective delivery mechanism for masculinizing young transgender men. *LGBT Health*. 2014;**1**(3): 165–167.
125. Spratt DI, Stewart I, Savage C, Craig W, Spack NP, Chandler DW, Spratt LV, Eimicke T, Olshan JS. Subcutaneous injection of testosterone is an effective and preferred alternative to intramuscular injection: demonstration in female-to-male transgender patients. *J Clin Endocrinol Metab*. 2017. doi:10.1210/jc.2017-00359
126. Eisenegger C, von Eckardstein A, Fehr E, von Eckardstein S. Pharmacokinetics of testosterone and estradiol gel preparations in healthy young men. *Psychoneuroendocrinology*. 2013;**38**(2): 171–178.
127. de Ronde W, ten Kulve J, Woerdeman J, Kaufman J-M, de Jong FH. Effects of oestradiol on gonadotrophin levels in normal and castrated men. *Clin Endocrinol (Oxf)*. 2009;**71**(6):874–879.
128. Money J, Ehrhardt A. Man & woman, boy & girl: differentiation and dimorphism of gender identity from conception to maturity. Baltimore, MD: Johns Hopkins University Press; 1972:202–206.
129. Heylens G, Verroken C, De Cock S, T'Sjoen G, De Cuypere G. Effects of different steps in gender reassignment therapy on psychopathology: a prospective study of persons with a gender identity disorder. *J Sex Med*. 2014;**11**(1):119–126.
130. Costa R, Colizzi M. The effect of cross-sex hormonal treatment on gender dysphoria individuals' mental health: a systematic review. *Neuropsychiatr Dis Treat*. 2016;**12**:1953–1966.
131. Gooren LJG, Giltay EJ. Review of studies of androgen treatment of female-to-male transsexuals: effects and risks of administration of androgens to females. *J Sex Med*. 2008;**5**(4):765–776.
132. Levy A, Crown A, Reid R. Endocrine intervention for transsexuals. *Clin Endocrinol (Oxf)*. 2003;**59**(4):409–418.
133. Tangpricha V, Ducharme SH, Barber TW, Chipkin SR. Endocrinologic treatment of gender identity disorders. *Endocr Pract*. 2003;**9**(1):12–21.
134. Meriggiola MC, Gava G. Endocrine care of transpeople part I. A review of cross-sex hormonal treatments, outcomes and adverse effects in transmen. *Clin Endocrinol (Oxf)*. 2015;**83**(5):597–606.
135. Bhasin S, Cunningham GR, Hayes FJ, Matsumoto AM, Snyder PJ, Swerdloff RS, Montori VM. Testosterone therapy in adult men with androgen deficiency syndromes: an endocrine society clinical practice guideline. *J Clin Endocrinol Metab*. 2006;**91**(6): 1995–2010.
136. Pelusi C, Costantino A, Martelli V, Lambertini M, Bazzocchi A, Ponti F, Battista G, Venturoli S, Meriggiola MC. Effects of three different testosterone formulations in female-to-male transsexual persons. *J Sex Med*. 2014;**11**(12):3002–3011.
137. Anderson GL, Limacher M, Assaf AR, Bassford T, Beresford SA, Black H, Bonds D, Brunner R, Brzyski R, Caan B, Chlebowski R, Curb D, Gass M, Hays J, Heiss G, Hendrix S, Howard BV, Hsia J, Hubbell A, Jackson R, Johnson KC, Judd H, Kotchen JM, Kuller L, LaCroix AZ, Lane D, Langer RD, Lasser N, Lewis CE, Manson J, Margolis K, Ockene J, O'Sullivan MJ, Phillips L, Prentice RL, Ritenbaugh C, Robbins J, Rossouw JE, Sarto G, Stefanick ML, Van Horn L, Wactawski-Wende J, Wallace R, Wassertheil-Smoller S; Women's Health Initiative Steering Committee. Effects of conjugated equine estrogen in postmenopausal women with hysterectomy: the Women's Health Initiative randomized controlled trial. *JAMA*. 2004;**291**(14):1701–1712.
138. Dickersin K, Munro MG, Clark M, Langenberg P, Scherer R, Frick K, Zhu Q, Hallock L, Nichols J, Yalcinkaya TM; Surgical Treatments Outcomes Project for Dysfunctional Uterine Bleeding (STOP-DUB) Research Group. Hysterectomy compared with endometrial ablation for dysfunctional uterine bleeding: a randomized controlled trial. *Obstet Gynecol*. 2007;**110**(6): 1279–1289.
139. Gooren LJ, Giltay EJ, Bunck MC. Long-term treatment of transsexuals with cross-sex hormones: extensive personal experience. *J Clin Endocrinol Metab*. 2008;**93**(1):19–25.
140. Prior JC, Vigna YM, Watson D. Spironolactone with physiological female steroids for presurgical therapy of male-to-female transsexualism. *Arch Sex Behav*. 1989;**18**(1):49–57.
141. Dittrich R, Binder H, Cupisti S, Hoffmann I, Beckmann MW, Mueller A. Endocrine treatment of male-to-female transsexuals using gonadotropin-releasing hormone agonist. *Exp Clin Endocrinol Diabetes*. 2005;**113**(10):586–592.

142. Stripp B, Taylor AA, Bartter FC, Gillette JR, Loriaux DL, Easley R, Menard RH. Effect of spironolactone on sex hormones in man. *J Clin Endocrinol Metab*. 1975;**41**(4):777–781.
143. Levy J, Burshell A, Marbach M, Afllalo L, Glick SM. Interaction of spironolactone with oestradiol receptors in cytosol. *J Endocrinol*. 1980;**84**(3):371–379.
144. Wierckx K, Elaut E, Van Hoorde B, Heylens G, De Cuypere G, Monstrey S, Weyers S, Hoebeke P, T'Sjoen G. Sexual desire in trans persons: associations with sex reassignment treatment. *J Sex Med*. 2014;**11**(1):107–118.
145. Chiriacò G, Cauci S, Mazzon G, Trombetta C. An observational retrospective evaluation of 79 young men with long-term adverse effects after use of finasteride against androgenetic alopecia. *Andrology*. 2016;**4**(2):245–250.
146. Gava G, Cerpolini S, Martelli V, Battista G, Seracchioli R, Meriggiola MC. Cyproterone acetate vs leuprolide acetate in combination with transdermal oestradiol in transwomen: a comparison of safety and effectiveness. *Clin Endocrinol (Oxf)*. 2016; **85**(2):239–246.
147. Casper RF, Yen SS. Rapid absorption of micronized estradiol-17 beta following sublingual administration. *Obstet Gynecol*. 1981; **57**(1):62–64.
148. Price TM, Blauer KL, Hansen M, Stanczyk F, Lobo R, Bates GW. Single-dose pharmacokinetics of sublingual versus oral administration of micronized 17β-estradiol. *Obstet Gynecol*. 1997;**89**(3): 340–345.
149. Toorians AWFT, Thomassen MCLGD, Zweegman S, Magdeleyns EJP, Tans G, Gooren LJG, Rosing J. Venous thrombosis and changes of hemostatic variables during cross-sex hormone treatment in transsexual people. *J Clin Endocrinol Metab*. 2003;**88**(12): 5723–5729.
150. Mepham N, Bouman WP, Arcelus J, Hayter M, Wylie KR. People with gender dysphoria who self-prescribe cross-sex hormones: prevalence, sources, and side effects knowledge. *J Sex Med*. 2014; **11**(12):2995–3001.
151. Richards C, Bouman WP, Seal L, Barker MJ, Nieder TO, T'Sjoen G. Non-binary or genderqueer genders. *Int Rev Psychiatry*. 2016; **28**(1):95–102.
152. Cosyns M, Van Borsel J, Wierckx K, Dedecker D, Van de Peer F, Daelman T, Laenen S, T'Sjoen G. Voice in female-to-male transsexual persons after long-term androgen therapy. *Laryngoscope*. 2014;**124**(6):1409–1414.
153. Deuster D, Matulat P, Knief A, Zitzmann M, Rosslau K, Szukaj M, am Zehnhoff-Dinnesen A, Schmidt CM. Voice deepening under testosterone treatment in female-to-male gender dysphoric individuals. *Eur Arch Otorhinolaryngol*. 2016;**273**(4):959–965.
154. Lapauw B, Taes Y, Simoens S, Van Caenegem E, Weyers S, Goemaere S, Toye K, Kaufman J-M, T'Sjoen GG. Body composition, volumetric and areal bone parameters in male-to-female transsexual persons. *Bone*. 2008;**43**(6):1016–1021.
155. Meyer III WJ, Webb A, Stuart CA, Finkelstein JW, Lawrence B, Walker PA. Physical and hormonal evaluation of transsexual patients: a longitudinal study. *Arch Sex Behav*. 1986;**15**(2): 121–138.
156. Asscheman H, Gooren LJ, Assies J, Smits JP, de Slegte R. Prolactin levels and pituitary enlargement in hormone-treated male-to-female transsexuals. *Clin Endocrinol (Oxf)*. 1988;**28**(6):583–588.
157. Gooren LJ, Harmsen-Louman W, van Kessel H. Follow-up of prolactin levels in long-term oestrogen-treated male-to-female transsexuals with regard to prolactinoma induction. *Clin Endocrinol (Oxf)*. 1985;**22**(2):201–207.
158. Wierckx K, Van Caenegem E, Schreiner T, Haraldsen I, Fisher AD, Toye K, Kaufman JM, T'Sjoen G. Cross-sex hormone therapy in trans persons is safe and effective at short-time follow-up: results from the European network for the investigation of gender incongruence. *J Sex Med*. 2014;**11**(8):1999–2011.
159. Ott J, Kaufmann U, Bentz EK, Huber JC, Tempfer CB. Incidence of thrombophilia and venous thrombosis in transsexuals under cross-sex hormone therapy. *Fertil Steril*. 2010;**93**(4):1267–1272.
160. Giltay EJ, Hoogeveen EK, Elbers JMH, Gooren LJG, Asscheman H, Stehouwer CDA. Effects of sex steroids on plasma total homocysteine levels: a study in transsexual males and females. *J Clin Endocrinol Metab*. 1998;**83**(2):550–553.
161. van Kesteren PJM, Asscheman H, Megens JAJ, Gooren LJG. Mortality and morbidity in transsexual subjects treated with cross-sex hormones. *Clin Endocrinol (Oxf)*. 1997;**47**(3): 337–343.
162. Wierckx K, Gooren L, T'Sjoen G. Clinical review: breast development in trans women receiving cross-sex hormones. *J Sex Med*. 2014;**11**(5):1240–1247.
163. Bird D, Vowles K, Anthony PP. Spontaneous rupture of a liver cell adenoma after long term methyltestosterone: report of a case successfully treated by emergency right hepatic lobectomy. *Br J Surg*. 1979;**66**(3):212–213.
164. Westaby D, Ogle SJ, Paradinas FJ, Randell JB, Murray-Lyon IM. Liver damage from long-term methyltestosterone. *Lancet*. 1977; **2**(8032):262–263.
165. Weinand JD, Safer JD. Hormone therapy in transgender adults is safe with provider supervision; a review of hormone therapy sequelae for transgender individuals. *J Clin Transl Endocrinol*. 2015;**2**(2):55–60.
166. Roberts TK, Kraft CS, French D, Ji W, Wu AH, Tangpricha V, Fantz CR. Interpreting laboratory results in transgender patients on hormone therapy. *Am J Med*. 2014;**127**(2):159–162.
167. Vesper HW, Botelho JC, Wang Y. Challenges and improvements in testosterone and estradiol testing. *Asian J Androl*. 2014;**16**(2): 178–184.
168. Asscheman H, T'Sjoen G, Lemaire A, Mas M, Meriggiola MC, Mueller A, Kuhn A, Dhejne C, Morel-Journel N, Gooren LJ. Venous thrombo-embolism as a complication of cross-sex hormone treatment of male-to-female transsexual subjects: a review. *Andrologia*. 2014;**46**(7):791–795.
169. Righini M, Perrier A, De Moerloose P, Bounameaux H. D-dimer for venous thromboembolism diagnosis: 20 years later. *J Thromb Haemost*. 2008;**6**(7):1059–1071.
170. Gooren LJ, Assies J, Asscheman H, de Slegte R, van Kessel H. Estrogen-induced prolactinoma in a man. *J Clin Endocrinol Metab*. 1988;**66**(2):444–446.
171. Kovacs K, Stefaneanu L, Ezzat S, Smyth HS. Prolactin-producing pituitary adenoma in a male-to-female transsexual patient with protracted estrogen administration. A morphologic study. *Arch Pathol Lab Med*. 1994;**118**(5):562–565.
172. Serri O, Noiseux D, Robert F, Hardy J. Lactotroph hyperplasia in an estrogen treated male-to-female transsexual patient. *J Clin Endocrinol Metab*. 1996;**81**(9):3177–3179.
173. Cunha FS, Domenice S, Câmara VL, Sircili MH, Gooren LJ, Mendonça BB, Costa EM. Diagnosis of prolactinoma in two male-to-female transsexual subjects following high-dose cross-sex hormone therapy. *Andrologia*. 2015;**47**(6):680–684.
174. Nota NM, Dekker MJHJ, Klaver M, Wiepjes CM, van Trotsenburg MA, Heijboer AC, den Heijer M. Prolactin levels during short- and long-term cross-sex hormone treatment: an observational study in transgender persons. *Andrologia*. 2017;**49**(6).
175. Bunck MC, Debono M, Giltay EJ, Verheijen AT, Diamant M, Gooren LJ. Autonomous prolactin secretion in two male-to-female transgender patients using conventional oestrogen dosages. *BMJ Case Rep*. 2009;**2009**:bcr0220091589.
176. Elamin MB, Garcia MZ, Murad MH, Erwin PJ, Montori VM. Effect of sex steroid use on cardiovascular risk in transsexual individuals: a systematic review and meta-analyses. *Clin Endocrinol (Oxf)*. 2010;**72**(1):1–10.
177. Berra M, Armillotta F, D'Emidio L, Costantino A, Martorana G, Pelusi G, Meriggiola MC. Testosterone decreases adiponectin

levels in female to male transsexuals. *Asian J Androl*. 2006;**8**(6):725–729.
178. Elbers JMH, Giltay EJ, Teerlink T, Scheffer PG, Asscheman H, Seidell JC, Gooren LJG. Effects of sex steroids on components of the insulin resistance syndrome in transsexual subjects. *Clin Endocrinol (Oxf)*. 2003;**58**(5):562–571.
179. Giltay EJ, Lambert J, Gooren LJG, Elbers JMH, Steyn M, Stehouwer CDA. Sex steroids, insulin, and arterial stiffness in women and men. *Hypertension*. 1999;**34**(4 Pt 1):590–597.
180. Polderman KH, Gooren LJ, Asscheman H, Bakker A, Heine RJ. Induction of insulin resistance by androgens and estrogens. *J Clin Endocrinol Metab*. 1994;**79**(1):265–271.
181. Maraka S. Effect of sex steroids on lipids, venous thromboembolism, cardiovascular disease and mortality in transgender individuals: a systematic review and meta-analysis. Available at: http://press.endocrine.org/doi/abs/10.1210/endo-meetings.2016.RE.15.FRI-136. Accessed 3 July 2017.
182. Meriggiola MC, Armillotta F, Costantino A, Altieri P, Saad F, Kalhorn T, Perrone AM, Ghi T, Pelusi C, Pelusi G. Effects of testosterone undecanoate administered alone or in combination with letrozole or dutasteride in female to male transsexuals. *J Sex Med*. 2008;**5**(10):2442–2453.
183. Giltay EJ, Toorians AW, Sarabdjitsingh AR, de Vries NA, Gooren LJ. Established risk factors for coronary heart disease are unrelated to androgen-induced baldness in female-to-male transsexuals. *J Endocrinol*. 2004;**180**(1):107–112.
184. Giltay EJ, Verhoef P, Gooren LJG, Geleijnse JM, Schouten EG, Stehouwer CDA. Oral and transdermal estrogens both lower plasma total homocysteine in male-to-female transsexuals. *Atherosclerosis*. 2003;**168**(1):139–146.
185. Calof OM, Singh AB, Lee ML, Kenny AM, Urban RJ, Tenover JL, Bhasin S. Adverse events associated with testosterone replacement in middle-aged and older men: a meta-analysis of randomized, placebo-controlled trials. *J Gerontol A Biol Sci Med Sci*. 2005;**60**(11):1451–1457.
186. Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults. Executive summary of the Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). *JAMA*. 2001;**285**(19):2486–2497.
187. Murad MH, Elamin MB, Garcia MZ, Mullan RJ, Murad A, Erwin PJ, Montori VM. Hormonal therapy and sex reassignment: a systematic review and meta-analysis of quality of life and psychosocial outcomes. *Clin Endocrinol (Oxf)*. 2010;**72**(2):214–231.
188. Van Caenegem E, Wierckx K, Taes Y, Schreiner T, Vandewalle S, Toye K, Lapauw B, Kaufman JM, T'Sjoen G. Body composition, bone turnover, and bone mass in trans men during testosterone treatment: 1-year follow-up data from a prospective case-controlled study (ENIGI). *Eur J Endocrinol*. 2015;**172**(2):163–171.
189. Turner A, Chen TC, Barber TW, Malabanan AO, Holick MF, Tangpricha V. Testosterone increases bone mineral density in female-to-male transsexuals: a case series of 15 subjects. *Clin Endocrinol (Oxf)*. 2004;**61**(5):560–566.
190. van Kesteren P, Lips P, Gooren LJG, Asscheman H, Megens J. Long-term follow-up of bone mineral density and bone metabolism in transsexuals treated with cross-sex hormones. *Clin Endocrinol (Oxf)*. 1998;**48**(3):347–354.
191. Van Caenegem E, Taes Y, Wierckx K, Vandewalle S, Toye K, Kaufman JM, Schreiner T, Haraldsen I, T'Sjoen G. Low bone mass is prevalent in male-to-female transsexual persons before the start of cross-sex hormonal therapy and gonadectomy. *Bone*. 2013;**54**(1):92–97.
192. Amin S, Zhang Y, Sawin CT, Evans SR, Hannan MT, Kiel DP, Wilson PW, Felson DT. Association of hypogonadism and estradiol levels with bone mineral density in elderly men from the Framingham study. *Ann Intern Med*. 2000;**133**(12):951–963.
193. Gennari L, Khosla S, Bilezikian JP. Estrogen and fracture risk in men. *J Bone Miner Res*. 2008;**23**(10):1548–1551.
194. Khosla S, Melton LJ III, Atkinson EJ, O'Fallon WM, Klee GG, Riggs BL. Relationship of serum sex steroid levels and bone turnover markers with bone mineral density in men and women: a key role for bioavailable estrogen. *J Clin Endocrinol Metab*. 1998;**83**(7):2266–2274.
195. Mueller A, Dittrich R, Binder H, Kuehnel W, Maltaris T, Hoffmann I, Beckmann MW. High dose estrogen treatment increases bone mineral density in male-to-female transsexuals receiving gonadotropin-releasing hormone agonist in the absence of testosterone. *Eur J Endocrinol*. 2005;**153**(1):107–113.
196. Ruetsche AG, Kneubuehl R, Birkhaeuser MH, Lippuner K. Cortical and trabecular bone mineral density in transsexuals after long-term cross-sex hormonal treatment: a cross-sectional study. *Osteoporos Int*. 2005;**16**(7):791–798.
197. Ganly I, Taylor EW. Breast cancer in a trans-sexual man receiving hormone replacement therapy. *Br J Surg*. 1995;**82**(3):341.
198. Pritchard TJ, Pankowsky DA, Crowe JP, Abdul-Karim FW. Breast cancer in a male-to-female transsexual. A case report. *JAMA*. 1988;**259**(15):2278–2280.
199. Symmers WS. Carcinoma of breast in trans-sexual individuals after surgical and hormonal interference with the primary and secondary sex characteristics. *BMJ*. 1968;**2**(5597):83–85.
200. Brown GR. Breast cancer in transgender veterans: a ten-case series. *LGBT Health*. 2015;**2**(1):77–80.
201. Shao T, Grossbard ML, Klein P. Breast cancer in female-to-male transsexuals: two cases with a review of physiology and management. *Clin Breast Cancer*. 2011;**11**(6):417–419.
202. Nikolic DV, Djordjevic ML, Granic M, Nikolic AT, Stanimirovic VV, Zdravkovic D, Jelic S. Importance of revealing a rare case of breast cancer in a female to male transsexual after bilateral mastectomy. *World J Surg Oncol*. 2012;**10**:280.
203. Bösze P, Tóth A, Török M. Hormone replacement and the risk of breast cancer in Turner's syndrome. *N Engl J Med*. 2006;**355**(24):2599–2600.
204. Schoemaker MJ, Swerdlow AJ, Higgins CD, Wright AF, Jacobs PA; UK Clinical Cytogenetics Group. Cancer incidence in women with Turner syndrome in Great Britain: a national cohort study. *Lancet Oncol*. 2008;**9**(3):239–246.
205. Smith RA, Cokkinides V, Eyre HJ. American Cancer Society guidelines for the early detection of cancer, 2006. *CA Cancer J Clin*. 2006;**56**(1):11–25, quiz 49–50.
206. Wilson JD, Roehrborn C. Long-term consequences of castration in men: lessons from the Skoptzy and the eunuchs of the Chinese and Ottoman courts. *J Clin Endocrinol Metab*. 1999;**84**(12):4324–4331.
207. van Kesteren P, Meinhardt W, van der Valk P, Geldof A, Megens J, Gooren L. Effects of estrogens only on the prostates of aging men. *J Urol*. 1996;**156**(4):1349–1353.
208. Brown JA, Wilson TM. Benign prostatic hyperplasia requiring transurethral resection of the prostate in a 60-year-old male-to-female transsexual. *Br J Urol*. 1997;**80**(6):956–957.
209. Casella R, Bubendorf L, Schaefer DJ, Bachmann A, Gasser TC, Sulser T. Does the prostate really need androgens to grow? Transurethral resection of the prostate in a male-to-female transsexual 25 years after sex-changing operation. *Urol Int*. 2005;**75**(3):288–290.
210. Dorff TB, Shazer RL, Nepomuceno EM, Tucker SJ. Successful treatment of metastatic androgen-independent prostate carcinoma in a transsexual patient. *Clin Genitourin Cancer*. 2007;**5**(5):344–346.
211. Thurston AV. Carcinoma of the prostate in a transsexual. *Br J Urol*. 1994;**73**(2):217.

212. van Harst EP, Newling DW, Gooren LJ, Asscheman H, Prenger DM. Metastatic prostatic carcinoma in a male-to-female transsexual. *BJU Int*. 1998;**81**:776.
213. Turo R, Jallad S, Prescott S, Cross WR. Metastatic prostate cancer in transsexual diagnosed after three decades of estrogen therapy. *Can Urol Assoc J*. 2013;**7**(7–8):E544–E546.
214. Miksad RA, Bubley G, Church P, Sanda M, Rofsky N, Kaplan I, Cooper A. Prostate cancer in a transgender woman 41 years after initiation of feminization. *JAMA*. 2006;**296**(19):2316–2317.
215. Moyer VA; U.S. Preventive Services Task Force. Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2012;**157**(2):120–134.
216. Futterweit W. Endocrine therapy of transsexualism and potential complications of long-term treatment. *Arch Sex Behav*. 1998; **27**(2):209–226.
217. Miller N, Bédard YC, Cooter NB, Shaul DL. Histological changes in the genital tract in transsexual women following androgen therapy. *Histopathology*. 1986;**10**(7):661–669.
218. O'Hanlan KA, Dibble SL, Young-Spint M. Total laparoscopic hysterectomy for female-to-male transsexuals. *Obstet Gynecol*. 2007;**110**(5):1096–1101.
219. Dizon DS, Tejada-Berges T, Koelliker S, Steinhoff M, Granai CO. Ovarian cancer associated with testosterone supplementation in a female-to-male transsexual patient. *Gynecol Obstet Invest*. 2006; **62**(4):226–228.
220. Hage JJ, Dekker JJML, Karim RB, Verheijen RHM, Bloemena E. Ovarian cancer in female-to-male transsexuals: report of two cases. *Gynecol Oncol*. 2000;**76**(3):413–415.
221. Mueller A, Gooren L. Hormone-related tumors in transsexuals receiving treatment with cross-sex hormones. *Eur J Endocrinol*. 2008;**159**(3):197–202.
222. Coleman E, Bockting W, Botzer M, Cohen-Kettenis P, DeCuypere G, Feldman J, Fraser L, Green J, Knudson G, Meyer WJ, Monstrey S, Adler RK, Brown GR, Devor AH, Ehrbar R, Ettner R, Eyler E, Garofalo R, Karasic DH, Lev AI, Mayer G, Meyer-Bahlburg H, Hall BP, Pfaefflin F, Rachlin K, Robinson B, Schechter LS, Tangpricha V, van Trotsenburg M, Vitale A, Winter S, Whittle S, Wylie KR, Zucker K. Standards of care for the health of transsexual, transgender, and gender-nonconforming people, version 7. *Int J Transgenderism*. 2012;**13**:165–232.
223. Colebunders B, D'Arpa S, Weijers S, Lumen N, Hoebeke P, Monstrey S. Female-to-male gender reassignment surgery. In: Ettner R, Monstrey S, Coleman E, eds. *Principles of Transgender Medicine and Surgery*. 2nd ed. New York, NY: Routledge Taylor & Francis Group; 2016:279–317.
224. Monstrey S, Hoebeke P, Dhont M, De Cuypere G, Rubens R, Moerman M, Hamdi M, Van Landuyt K, Blondeel P. Surgical therapy in transsexual patients: a multi-disciplinary approach. *Acta Chir Belg*. 2001;**101**(5):200–209.
225. Selvaggi G, Ceulemans P, De Cuypere G, VanLanduyt K, Blondeel P, Hamdi M, Bowman C, Monstrey S. Gender identity disorder: general overview and surgical treatment for vaginoplasty in male-to-female transsexuals. *Plast Reconstr Surg*. 2005;**116**(6): 135e–145e.
226. Tugnet N, Goddard JC, Vickery RM, Khoosal D, Terry TR. Current management of male-to-female gender identity disorder in the UK. *Postgrad Med J*. 2007;**83**(984):638–642.
227. Horbach SER, Bouman M-B, Smit JM, Özer M, Buncamper ME, Mullender MG. Outcome of vaginoplasty in male-to-female transgenders: a systematic review of surgical techniques. *J Sex Med*. 2015;**12**(6):1499–1512.
228. Wroblewski P, Gustafsson J, Selvaggi G. Sex reassignment surgery for transsexuals. *Curr Opin Endocrinol Diabetes Obes*. 2013; **20**(6):570–574.
229. Morrison SD, Satterwhite T, Grant DW, Kirby J, Laub DR, Sr, VanMaasdam J. Long-term outcomes of rectosigmoid neocolporrhaphy in male-to-female gender reassignment surgery. *Plast Reconstr Surg*. 2015;**136**(2):386–394.
230. Dessy LA, Mazzocchi M, Corrias F, Ceccarelli S, Marchese C, Scuderi N. The use of cultured autologous oral epithelial cells for vaginoplasty in male-to-female transsexuals: a feasibility, safety, and advantageousness clinical pilot study. *Plast Reconstr Surg*. 2014;**133**(1):158–161.
231. Li FY, Xu YS, Zhou CD, Zhou Y, Li SK, Li Q. Long-term outcomes of vaginoplasty with autologous buccal micromucosa. *Obstet Gynecol*. 2014;**123**(5):951–956.
232. Kanhai RC. Sensate vagina pedicled-spot for male-to-female transsexuals: the experience in the first 50 patients. *Aesthetic Plast Surg*. 2016;**40**(2):284–287.
233. Straayer C. Transplants for transsexuals? Ambitions, concerns, ideology. Paper presented at: Trans*Studies: An International Transdisciplinary Conference on Gender, Embodiment, and Sexuality; 7–10 September 2016; University of Arizona, Tucson, AZ.
234. Bucci S, Mazzon G, Liguori G, Napoli R, Pavan N, Bormioli S, Ollandini G, De Concilio B, Trombetta C. Neovaginal prolapse in male-to-female transsexuals: an 18-year-long experience. *Biomed Res Int*. 2014;**2014**:240761.
235. Raigosa M, Avvedimento S, Yoon TS, Cruz-Gimeno J, Rodriguez G, Fontdevila J. Male-to-female genital reassignment surgery: a retrospective review of surgical technique and complications in 60 patients. *J Sex Med*. 2015;**12**(8):1837–1845.
236. Green R. Sexual functioning in post-operative transsexuals: male-to-female and female-to-male. *Int J Impot Res*. 1998;**10**(Suppl 1): S22–S24.
237. Hess J, Rossi Neto R, Panic L, Rübben H, Senf W. Satisfaction with male-to-female gender reassignment surgery. *Dtsch Arztebl Int*. 2014;**111**(47):795–801.
238. Nygren U, Nordenskjold A, Arver S, Sodersten M. Effects on voice fundamental frequency and satisfaction with voice in trans men during testosterone treatment—a longitudinal study. *J Voice*. 2016;**30**(6):766.e23-766.e34.
239. Becking AG, Tuinzing DB, Hage JJ, Gooren LJG. Transgender feminization of the facial skeleton. *Clin Plast Surg*. 2007;**34**(3): 557–564.
240. Giraldo F, Esteva I, Bergero T, Cano G, González C, Salinas P, Rivada E, Lara JS, Soriguer F; Andalusia Gender Team. Corona glans clitoroplasty and urethropreputial vestibuloplasty in male-to-female transsexuals: the vulval aesthetic refinement by the Andalusia Gender Team. *Plast Reconstr Surg*. 2004;**114**(6): 1543–1550.
241. Goddard JC, Vickery RM, Terry TR. Development of feminizing genitoplasty for gender dysphoria. *J Sex Med*. 2007;**4**(4 Pt 1): 981–989.
242. Hage JJ, de Graaf FH, Bouman FG, Bloem JJAM. Sculpturing the glans in phalloplasty. *Plast Reconstr Surg*. 1993;**92**(1):157–161, discussion 162.
243. Thiagaraj D, Gunasegaram R, Loganath A, Peh KL, Kottegoda SR, Ratnam SS. Histopathology of the testes from male transsexuals on oestrogen therapy. *Ann Acad Med Singapore*. 1987; **16**(2):347–348.
244. Monstrey SJ, Ceulemans P, Hoebeke P. Sex reassignment surgery in the female-to-male transsexual. *Semin Plast Surg*. 2011;**25**(3): 229–244.
245. Perovic SV, Djinovic R, Bumbasirevic M, Djordjevic M, Vukovic P. Total phalloplasty using a musculocutaneous latissimus dorsi flap. *BJU Int*. 2007;**100**(4):899–905, discussion 905.
246. Vesely J, Hyza P, Ranno R, Cigna E, Monni N, Stupka I, Justan I, Dvorak Z, Novak P, Ranno S. New technique of total phalloplasty with reinnervated latissimus dorsi myocutaneous free flap in female-to-male transsexuals. *Ann Plast Surg*. 2007;**58**(5): 544–550.
247. Ranno R, Veselý J, Hýza P, Stupka I, Justan I, Dvorák Z, Monni N, Novák P, Ranno S. Neo-phalloplasty with re-innervated latissimus dorsi free flap: a functional study of a novel technique. *Acta Chir Plast*. 2007;**49**(1):3–7.

248. Garcia MM, Christopher NA, De Luca F, Spilotros M, Ralph DJ. Overall satisfaction, sexual function, and the durability of neophallus dimensions following staged female to male genital gender confirming surgery: the Institute of Urology, London U.K. experience. *Transl Androl Urol*. 2014;**3**(2):156–162.
249. Chen H-C, Gedebou TM, Yazar S, Tang Y-B. Prefabrication of the free fibula osteocutaneous flap to create a functional human penis using a controlled fistula method. *J Reconstr Microsurg*. 2007;**23**(3):151–154.
250. Hoebeke PB, Decaestecker K, Beysens M, Opdenakker Y, Lumen N, Monstrey SM. Erectile implants in female-to-male transsexuals: our experience in 129 patients. *Eur Urol*. 2010;**57**(2):334–341.
251. Hage JJ. Metaidoioplasty: an alternative phalloplasty technique in transsexuals. *Plast Reconstr Surg*. 1996;**97**(1):161–167.
252. Cohanzad S. Extensive metoidioplasty as a technique capable of creating a compatible analogue to a natural penis in female transsexuals. *Aesthetic Plast Surg*. 2016;**40**(1):130–138.
253. Selvaggi G, Hoebeke P, Ceulemans P, Hamdi M, Van Landuyt K, Blondeel P, De Cuypere G, Monstrey S. Scrotal reconstruction in female-to-male transsexuals: a novel scrotoplasty. *Plast Reconstr Surg*. 2009;**123**(6):1710–1718.
254. Bjerrome Ahlin H, Kölby L, Elander A, Selvaggi G. Improved results after implementation of the Ghent algorithm for subcutaneous mastectomy in female-to-male transsexuals. *J Plast Surg Hand Surg*. 2014;**48**(6):362–367.
255. Wolter A, Diedrichson J, Scholz T, Arens-Landwehr A, Liebau J. Sexual reassignment surgery in female-to-male transsexuals: an algorithm for subcutaneous mastectomy. *J Plast Reconstr Aesthet Surg*. 2015;**68**(2):184–191.
256. Richards C, Barrett J. The case for bilateral mastectomy and male chest contouring for the female-to-male transsexual. *Ann R Coll Surg Engl*. 2013;**95**(2):93–95.
257. Sutcliffe PA, Dixon S, Akehurst RL, Wilkinson A, Shippam A, White S, Richards R, Caddy CM. Evaluation of surgical procedures for sex reassignment: a systematic review. *J Plast Reconstr Aesthet Surg*. 2009;**62**(3):294–306, discussion 306–308.
258. Selvaggi G, Elander A. Penile reconstruction/formation. *Curr Opin Urol*. 2008;**18**(6):589–597.
259. Dhejne C, Lichtenstein P, Boman M, Johansson ALV, Långström N, Landén M. Long-term follow-up of transsexual persons undergoing sex reassignment surgery: cohort study in Sweden. *PLoS One*. 2011;**6**(2):e16885.
260. Kuhn A, Bodmer C, Stadlmayr W, Kuhn P, Mueller MD, Birkhäuser M. Quality of life 15 years after sex reassignment surgery for transsexualism. *Fertil Steril*. 2009;**92**(5):1685–1689.e3.
261. Papadopulos NA, Lellé JD, Zavlin D, Herschbach P, Henrich G, Kovacs L, Ehrenberger B, Kluger AK, Machens HG, Schaff J. Quality of life and patient satisfaction following male-to-female sex reassignment surgery. *J Sex Med*. 2017;**14**(5):721–730.
262. Simonsen RK, Hald GM, Kristensen E, Giraldi A. Long-term follow-up of individuals undergoing sex-reassignment surgery: somatic morbidity and cause of death. *Sex Med*. 2016;**4**(1):e60–e68.
263. Djordjevic ML, Bizic MR, Duisin D, Bouman MB, Buncamper M. Reversal Surgery in regretful male-to-female transsexuals after sex reassignment surgery. *J Sex Med*. 2016;**13**(6):1000–1007.
264. Liberopoulos EN, Florentin M, Mikhailidis DP, Elisaf MS. Compliance with lipid-lowering therapy and its impact on cardiovascular morbidity and mortality. *Expert Opin Drug Saf*. 2008;**7**(6):717–725.
265. Forbes SS, Stephen WJ, Harper WL, Loeb M, Smith R, Christoffersen EP, McLean RF. Implementation of evidence-based practices for surgical site infection prophylaxis: results of a pre- and postintervention study. *J Am Coll Surg*. 2008;**207**(3):336–341.
266. Davis PJ, Spady D, de Gara C, Forgie SE. Practices and attitudes of surgeons toward the prevention of surgical site infections: a provincial survey in Alberta, Canada. *Infect Control Hosp Epidemiol*. 2008;**29**(12):1164–1166.

IN THE UNITED STATES DISTRICT COURT
FOR THE MIDDLE DISTRICT OF ALABAMA
NORTHERN DIVISION

* * *

BRIANNA BOE, et al.,
Plaintiffs,
UNITED STATES OF AMERICA,
Intervenor Plaintiff,
vs. CASE NO. 2:22-cv-184-LCB
HON. STEVE MARSHALL, in his Official capacity as Attorney General, of the State of Alabama, et al.,

Defendants.

* * *

Deposition of ARMAND H. ANTOMMARIA, M.D., Ph.D., FAAP, HEC-C, Witness herein, called by the Defendants for examination pursuant to the Rules of Civil Procedure, taken before me, Monica K. Schrader, a Notary Public in and for the State of Ohio, at the U.S. Attorney's Office, Cleveland Branch Office, Atrium II Building, 221 East Fourth Street, Suite 400, Cincinnati, Ohio, on Friday, April 21, 2023, at 9:03 a.m.

* * *

Page 206

the condition, correct? 15:02:23
A. If medical therapy was 15:02:23
unsuccessful, surgery might be considered, sir. 15:02:28
Q. And you can have with that 15:02:30
condition emergency situations that require 15:02:34
surgery, correct, like a bleed or perforation, 15:02:36
if you know? 15:02:42
A. I don't know that surgery would be 15:02:44
necessarily the primary intervention for 15:02:46
bleeding, but for perforation, yes, sir. 15:02:48
Q. Because if a perforation is left 15:02:52
untreated, that can cause death presumably, 15:02:54
right? 15:02:57
A. It can cause peritonitis, which 15:02:58
would be an infection in the abdominal cavity 15:03:01
which if left untreated could result in death, 15:03:06
sir. 15:03:08
Q. For a natal male at Tanner Stage 2 15:03:09
seeking to begin puberty blockers, what are the 15:03:22
options for preserving that child's fertility? 15:03:26
A. The primary option for preserving 15:03:29
fertility in that case would be delaying the 15:03:38
use of puberty blockers, sir. 15:03:41
Q. So you wouldn't actually start 15:03:43
them at Tanner 2 if you were trying to preserve 15:03:45

Page 207

fertility? 15:03:48
MR. CHEEK: Objection, foundation. 15:03:48
THE WITNESS: If that was your 15:03:50
exclusive or predominant goal, there would be a 15:03:56
reason to delay utilizing puberty blockers. There 15:04:00
might be other ways later in the future that by 15:04:05
discontinuing gender-affirming medical care 15:04:13
fertility could be reestablished. 15:04:16
BY MR. FRAMPTON: 15:04:20
Q. Have you seen any studies showing 15:04:22
the success of that process? 15:04:23
A. I am aware of studies that show 15:04:28
the resumption of fertility in individuals who 15:04:34
have discontinued gender-affirming hormone 15:04:37
therapy, sir. 15:04:41
Q. Aware of any studies dealing with 15:04:41
individuals who started puberty suppression at 15:04:44
Tanner Stage 2? 15:04:47
MR. CHEEK: Objection, form. 15:04:48
THE WITNESS: Not specifically of 15:04:50
that population, sir. 15:04:53
BY MR. FRAMPTON: 15:04:54
Q. Just as a general matter, if a 15:04:54
natal male starts puberty suppression at Tanner 15:04:57
Stage 2, continues seamlessly into estrogen and 15:05:01

Page 208

anti-androgen therapy, that person will never 15:05:06
develop fertility, correct, without stopping 15:05:10
treatment? 15:05:14
A. So, in general, the expectation 15:05:14
would be if that individual continued 15:05:19
treatment, that is correct that they would not 15:05:23
be fertile. 15:05:25
Q. And, likewise, with a natal female 15:05:26
who begins puberty suppression at Tanner Stage 15:05:30
2 and progresses seamlessly to testosterone 15:05:34
therapy, that individual would not develop 15:05:38
fertility, correct? 15:05:41
A. If they continued on treatment, 15:05:43
they would not be anticipated to have 15:05:51
biologically related children. It is to say 15:05:53
that for some individuals the benefit of 15:05:56
treatment would outweigh that risk, but that 15:05:59
risk would exist. 15:06:01
Q. And it wouldn't be a risk, it 15:06:02
would be they are not going to have fertility 15:06:12
without discontinuing treatment, correct? 15:06:15
MR. CHEEK: Objection, form. 15:06:20
THE WITNESS: I'm sorry, I don't 15:06:21
understand the distinction that you are making, 15:06:22
sir. 15:06:24

Page 209

BY MR. FRAMPTON: 15:06:24
Q. Well, I think you were 15:06:25
characterizing it as a risk of infertility, and 15:06:26
I was distinguishing it's really -- without 15:06:30
discontinuing treatment, it's a certainty of 15:06:33
infertility, is it not? 15:06:36
A. So when -- as an emphasis, when I 15:06:37
would refer to a risk, I wouldn't say that 15:06:40
risks involve both a magnitude and a 15:06:42
probability. So while colloquially risk might 15:06:44
have implications about probability, I don't 15:06:48
know that in the way an ethicist uses a risk 15:06:52
that it necessarily has those similar 15:06:57
implications. 15:07:04
Q. But you would agree, again, for 15:07:06
the natal female starting puberty suppression 15:07:08
at Tanner Stage 2 continuing seamlessly through 15:07:10
to testosterone therapy that that person -- you 15:07:13
would not have any expectation that person 15:07:16
would develop fertility with that course of 15:07:18
treatment, correct? 15:07:22
A. So given the currently available 15:07:22
resources for fertility preservation, no. 15:07:34
Q. Are you aware of any studies that 15:07:37
document healthy conception and birth by a 15:08:02