



2024 Health Econ Area of Emphasis Preliminary Exam

SCHOOL OF
PUBLIC HEALTH

UNIVERSITY OF MINNESOTA

How to Catch AI Usage

AI generated writing has common themes that once you notice are simple to catch:

Lack of personal experience included in the assignment; especially written papers

Inconsistent writing styles to previously turned in work

Unverifiable data, quotes, citations, or characters

Unusually complex sentences and word choices

You can use ZeroGPT and other checkers to confirm suspicions

AI prefers to write in numbered or bullet pointed lists

AI has “favorite” words and phrases that immediately raise red flags

5

Signs of LLM Use in Haishan Yang's Exam

- Unusually long answer, compared to recent prelims.
- Writing style is very structured (i.e., uses headings, subheadings, and bullet points throughout).
- Uses common phrasing for LLMs. Two instances of 'in summary' and one of 'in conclusion'.
- Noticeably identical or very similar text, when compared to ChatGPT output.
- Content that is non-standard for health economics (i.e., not in any of the papers on our prelim exam reading list), but does appear in the ChatGPT output
- Feels "voiceless" and does not read similarly to known examples of writing by this student.

Unusually Long Answer

In 2020, two students took our preliminary exam. Their word counts were 2482 and 4418.

In 2022, five students took our prelim. Their word counts 1691, 1756, 2853, 3011, and 3165.

In 2024, only Haishan Yang took our prelim and submitted an exam with 5570 words.

Haishan's exam was twice as long as the average exam. It was 26% longer than second-longest exam.

Similarities between Haishan Yang's submitted exam and ChatGPT output

Circumstances Leading to Suboptimal Service Levels

1. **Insufficient Risk Adjustment:** If the risk adjustment formula fails to accurately predict healthcare needs, providers may receive inadequate compensation for high-risk patients, leading to under-provision of services.
2. **Incentives for Cost Minimization:** Providers under capitation might focus excessively on reducing costs, potentially at the expense of necessary care. This could result in patients not receiving timely or appropriate treatments, adversely affecting health outcomes.

Comparing Haishan's answer to ChatGPT output, we observe:

An identical heading

Identical answer structure

Nearly identical answer content

ChatGPT Output:

Circumstances Leading to Suboptimal Service Levels: The level of services under global capitation could fall below what is socially optimal for patient health if:

1. **Inadequate Risk Adjustment:** If the capitation payments do not sufficiently account for the health status of patients, primary care organizations may be incentivized to reduce the quantity or intensity of services to maintain profitability, potentially compromising patient care.
2. **Excessive Cost-Cutting Pressures:** Organizations may seek to maximize their financial margins by cutting costs excessively, leading to under-provision of necessary services. This could happen if there are insufficient safeguards to ensure that patients receive appropriate care or if quality measures are not effectively enforced.

Haishan Yang's Answer to Question 4, Part B:

1.1 Risk Selection (Cream Skimming):

Under unconstrained competition, primary care organizations (PCOs) may engage in risk selection, also known as "cream skimming." This occurs when PCOs selectively enroll healthier patients who are less likely to require expensive healthcare services. By focusing on lower-risk individuals, these organizations can maximize their profits, as the fixed per-patient payments will likely exceed the cost of care for these patients. Conversely, sicker patients, who need more intensive and costly care, may find it difficult to enroll in these plans, leading to disparities in access to care and potential adverse health outcomes for higher-risk populations.

ChatGPT Output:

Circumstances Leading to Suboptimal Service Levels: The level of services under global capitation could fall below what is socially optimal for patient health if:

1. **Inadequate Risk Adjustment:** If the capitation payments do not sufficiently account for the health status of patients, primary care organizations may be incentivized to reduce the quantity or intensity of services to maintain profitability, potentially compromising patient care.
2. **Excessive Cost-Cutting Pressures:** Organizations may seek to maximize their financial margins by cutting costs excessively, leading to under-provision of necessary services. This could happen if there are insufficient safeguards to ensure that patients receive appropriate care or if quality measures are not effectively enforced.
3. **Patient Selection (Cream Skimming):** If primary care organizations are able to selectively enroll healthier patients (cream skimming) while avoiding sicker, more expensive patients, the overall level of service provision could be reduced, leading to inequities in care and potentially worse health outcomes for the sickest patients.

Comparing Haishan's answer to ChatGPT output, we observe:

Multiple instances of identical or near identical phrasing.

Another potential issue with unconstrained competition under global capitation is the under-provision of care. Since PCOs receive a fixed payment per patient, there is a financial incentive to minimize the quantity of services provided to increase profitability. This can lead to inadequate care, where patients do not receive the necessary medical interventions, follow-ups, or preventive services. Over time, under-provision of care can result in worsening health outcomes and higher long-term healthcare costs due to the progression of untreated conditions.

Comparing Haishan's answer to ChatGPT output, we observe:

Identical subheadings

Identical content

Implementation:

- **Regular Audits:** Conducting regular audits by independent bodies can verify that PCOs are meeting established care standards. These audits can assess various aspects of care, such as adherence to clinical guidelines, patient outcomes, and overall care quality.
- **Patient Satisfaction Surveys:** Gathering feedback directly from patients through regular surveys helps measure patient satisfaction and experience. High levels of patient satisfaction typically correlate with better care quality and patient outcomes.

ChatGPT Output

2. Under-provision of Care:

- **Problem:** PCOs might reduce the quantity or quality of care provided to patients to cut costs, leading to under-provision of necessary services. This could result in poorer health outcomes for patients, particularly those with chronic conditions or complex healthcare needs.
- **Regulatory Solution:** Establish minimum care standards and robust quality monitoring systems to ensure that all PCOs meet certain benchmarks for patient care. This could include regular audits, patient satisfaction surveys, and penalties for organizations that fail to meet quality standards.

Haishan Yang's Answer to Question 4, Part B:

Challenges and Solutions:

- **Data Accuracy:** Ensuring accurate and comprehensive data collection is crucial for effective risk adjustment. This can be achieved through standardized reporting systems and electronic health records.
- **Complexity:** Risk adjustment models can become complex, requiring sophisticated statistical methods and computational resources. Investing in robust healthcare IT infrastructure and analytical capabilities can address this complexity.

1.3 Minimum Care Requirements: Establishing minimum care requirements and standards can prevent PCOs from underserving high-risk patients. These requirements can include mandatory coverage of essential health services and quality benchmarks that PCOs must meet.

ChatGPT Output:

2. Under-provision of Care:

- o **Problem:** PCOs might reduce the quantity or quality of care provided to patients to cut costs, leading to under-provision of necessary services. This could result in poorer health outcomes for patients, particularly those with chronic conditions or complex healthcare needs.
- o **Regulatory Solution:** Establish minimum care standards and robust quality monitoring systems to ensure that all PCOs meet certain benchmarks for patient care. This could include regular audits, patient satisfaction surveys, and penalties for organizations that fail to meet quality standards.

Comparing Haishan's answer to ChatGPT output, we observe:

Identical phrasing

Weaknesses:

- **Compliance and Enforcement Issues:** Ensuring that individuals comply with the mandate can be challenging. The penalty may not be sufficiently high to compel all individuals to purchase insurance.
- **Equity Concerns:** The mandate might disproportionately affect lower-income individuals who might find it financially burdensome to purchase insurance, even with subsidies.
- **Political and Public Resistance:** Mandates can be politically contentious and may face significant public opposition. The Massachusetts mandate faced legal challenges, and similar opposition was observed with the ACA's individual mandate.

Comparing Haishan's answer to ChatGPT output, we observe:

Multiple instances of identical structure and phrasing.

ChatGPT Output:

Weaknesses:

- **Compliance and Enforcement:** The effectiveness of mandated coverage depends on the enforcement mechanism. If penalties for non-compliance are weak or difficult to enforce, the policy may fail to achieve its objective.
- **Equity Concerns:** Mandated coverage can be seen as inequitable, especially if low-income individuals are required to purchase insurance that they may not be able to afford, even with subsidies.

Haishan Yang's Answer to Question 3, Part A:

services is zero, leading to a higher quantity demanded than what would be optimal if they were paying out-of-pocket. Empirical evidence corroborates these theoretical predictions. The RAND Health Insurance Experiment found that individuals with more generous insurance coverage consumed significantly more healthcare services without corresponding improvements in health outcomes, indicating inefficient overuse of healthcare resources (Aron-Dine, Einav, and Finkelstein, 2013). However, it should be noted that the debate about the impact of RAND HIE is still ongoing, and there might be long-term health benefits. If there is an improvement in long-term health benefit, we should not view it as an example of moral hazard.

Comparing Haishan's answer to ChatGPT output, we observe:

Nearly identical phrasing

ChatGPT Output:

healthcare system. Studies have shown that when individuals face lower out-of-pocket costs for healthcare services, utilization increases significantly, even for services of marginal benefit. For example, the RAND Health Insurance Experiment, a landmark study, demonstrated that individuals with comprehensive health insurance used more healthcare services than those with higher cost-sharing, without a corresponding improvement in health outcomes. This evidence underscores the economic inefficiency generated by moral hazard.

Similarities between Haishan Yang's practice exam and ChatGPT output

Haishan Yang's Answer to the Practice Question:

Comparing Haishan's answer to ChatGPT output, we observe:

Multiple instances of nearly identical phrasing

Definition: Pure capitation is a payment arrangement in which health care providers or insurers receive a fixed amount of money per enrolled individual per period (e.g., per month), regardless of how many services the individual uses. This payment method is used to cover the costs of all or most of the services that an individual may need.

Incentives Created:

Cost Control: Providers or insurers have a strong incentive to control costs since their revenue is fixed and does not depend on the number or cost of services provided.

Preventive Care: There is an incentive to focus on preventive care and early intervention to avoid costly treatments later on.

Efficiency: Encourages efficiency and reduction of unnecessary services since providers bear the financial risk.

Risk Selection: May lead to risk selection, where providers might avoid enrolling individuals who are expected to have high healthcare costs.

ChatGPT Output:

Description: Pure capitation is a payment model in which a healthcare provider or insurer receives a fixed amount of money per patient per unit of time (e.g., per month), regardless of the number or type of services provided. This fixed amount is agreed upon in advance and is intended to cover all necessary services for the patient during that time period.

Incentives Created by Pure Capitation:

- Cost Control:** Providers are incentivized to reduce unnecessary services because they receive a fixed payment regardless of the number of services provided. This can lead to more efficient care delivery, focusing on preventive care and cost-effective treatments.
- Risk of Under-provision of Care:** Since providers are paid a fixed amount regardless of service usage, they may have an incentive to minimize care to reduce costs, potentially leading to under-provision of necessary services.
- Focus on Preventive Care:** Providers may emphasize preventive care to avoid the need for more expensive treatments, as they bear the financial risk of their patients' healthcare needs.
- Patient Selection:** Providers may be incentivized to avoid high-risk or high-cost patients, as these patients could cost more to treat than the capitation payment covers.

Content that is non-standard for doctoral health economics, but also appears in ChatGPT output

Haishan Yang's Answer to Question 4, Part B:

1.1 Risk Selection (Cream Skimming):

Under unconstrained competition, primary care organizations (PCOs) may engage in risk selection, also known as "cream skimming." This occurs when PCOs selectively enroll healthier patients who are less likely to require expensive healthcare services. By focusing on lower-risk individuals, these organizations can maximize their profits, as the fixed per-patient payments will likely exceed the cost of care for these patients. Conversely, sicker patients, who need more intensive and costly care, may find it difficult to enroll in these plans, leading to disparities in access to care and potential adverse health outcomes for higher-risk populations.

ChatGPT Output:

1. Risk Selection (Cream Skimming):

- **Problem:** PCOs might engage in risk selection, trying to attract healthier patients who are less likely to require expensive care. This can lead to a concentration of high-risk, high-cost patients in certain organizations, undermining the effectiveness of risk pooling and potentially destabilizing the market.

Haishan Yang's exam uses an acronym "PCO" that is not standard for health economics.

None of the four faculty members (all health economists) grading this prelim exam have seen this acronym used.

This acronym is not used in any papers included in the prelim exam reading list.

This acronym appears in the ChatGPT's output.

Haishan Yang's Answer to Question 4, Part B:

Implementation:

- **Regular Audits:** Conducting regular audits by independent bodies can verify that PCOs are meeting established care standards. These audits can assess various aspects of care, such as adherence to clinical guidelines, patient outcomes, and overall care quality.
- **Patient Satisfaction Surveys:** Gathering feedback directly from patients through regular surveys helps measure patient satisfaction and experience. High levels of patient satisfaction typically correlate with better care quality and patient outcomes.
- **Outcome Measures:** Tracking specific outcome measures, such as rates of hospital readmissions, control of chronic conditions, and preventive care utilization, provides concrete data on the effectiveness of the care provided. These measures can be aggregated and analyzed to identify trends and areas needing improvement.

Haishan Yang's exam proposes two regulatory solutions that are identical to those offered by ChatGPT, but are not standard for the field of health economics and are not discussed in any papers on the prelim exam reading list.

ChatGPT Output:

- **Regulatory Solution:** Establish minimum care standards and robust quality monitoring systems to ensure that all PCOs meet certain benchmarks for patient care. This could include regular audits, patient satisfaction surveys, and penalties for organizations that fail to meet quality standards.

GPTZero (an AI detector) Results

Our decision to include GPTZero results

We acknowledge that it is impossible to detect LLM-generated text with perfect accuracy.

- AI detectors such as GPTZero are known to sometimes make type 1 (false positives) and type 2 (false negatives) errors.

We include the following results only because we are able to compare Haishan Yang's prelim exam to a previous writing sample *known* to be his writing (i.e., a final exam that Haishan Yang hand-wrote for PubH 6832).

GPTZero assesses the probability of AI generation at 0% for Haishan Yang's known writing sample.

GPTZero assesses the probability of AI generation between 19% and 89% for Haishan Yang's answers to the prelim exam.

GPTZero Results for *Known Writing by Haishan Yang*

Final exam
From PubH
6832
(Fall 2022)

AI Classification

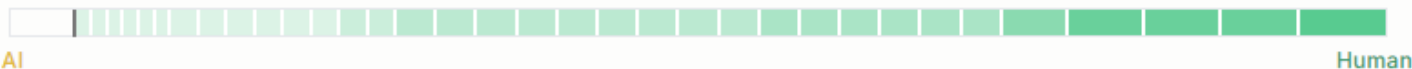


We are **highly confident** this text is entirely human.

0% Probability AI generated

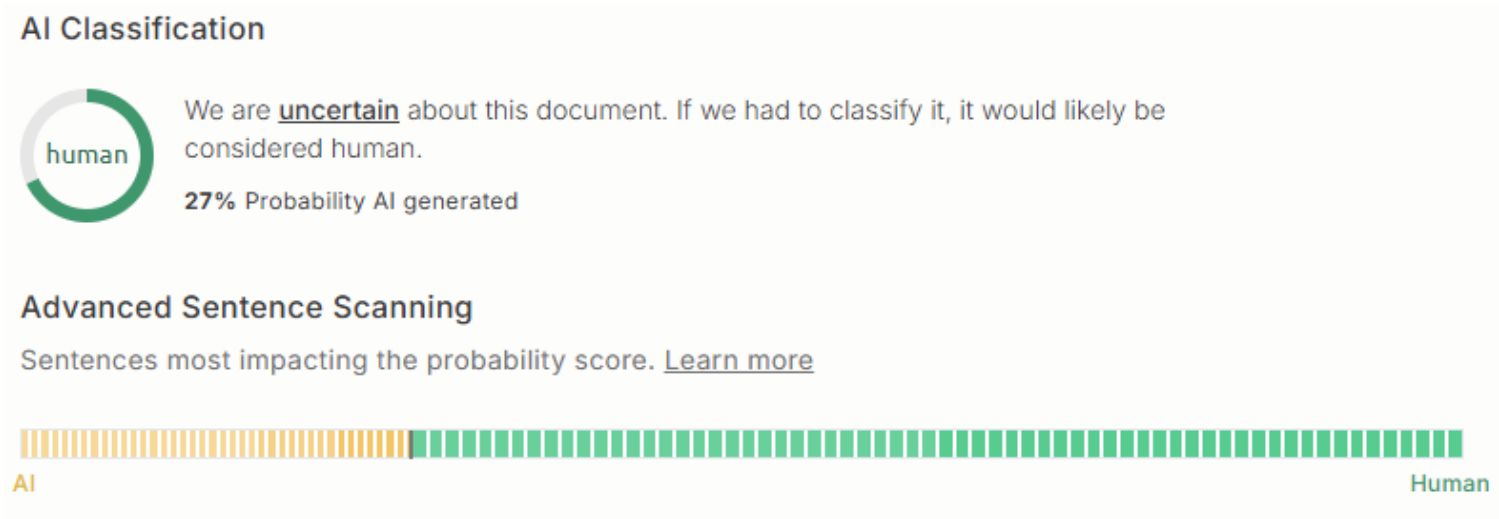
Advanced Sentence Scanning

Sentences most impacting the probability score. [Learn more](#)

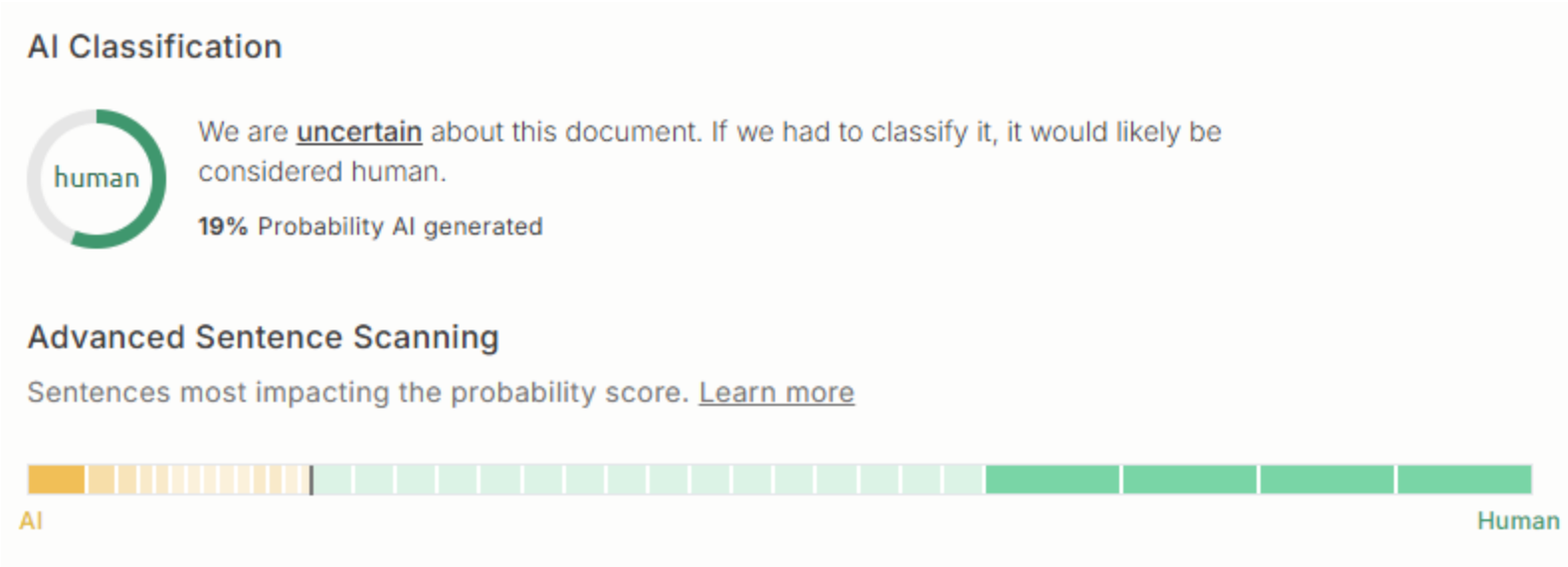


GPTZero Results for Haishan Yang's Prelim Submission

Q1



Q3



GPTZero Results for Haishan Yang's Prelim Submission

Q4

