

Concerns about Haishan Yang’s doctoral preliminary exam

1. The student took the health economics area-of-emphasis (AOE) written preliminary examination on August 5th, 2024. The preface to the exam stated, “The exam is open book (i.e., you may refer to lecture notes, papers, and textbooks). Please complete the exam individually. You may not consult with anyone else on the questions or content. You may not use any sort of Artificial intelligence tools, such as ChatGPT, in any part of this exam.” The format of the exam was that Question 1 was mandatory, and the student then chooses two of the subsequent four questions, so that they write answers for three questions altogether.

Of the four health economics AOE faculty that reviewed the exam, all four shared the initial reaction that the exam did not seem to be written in the student’s voice and noted instances where answers seemed to not be directly relevant to the question, or involved concepts that were not covered in class or the readings. All four AOE faculty had significant concerns that some sort of large language model/generative AI model was used in the answers.

2. In response to these concerns, we ran the three exam questions that were answered through ChatGPT. The ChatGPT output (included as an attachment) led to further concerns because of several specific instances of language and/or structure of the student’s answers matching the ChatGPT output closely. Here are a number of specific examples (labeled here as Q1, Q3, and Q4, per their labeling on the exam):

<p>Q1. A. Using the Rothchild-Stiglitz framework, describe the types of equilibria that can occur in insurance markets where consumers have heterogeneous risk types, under different assumptions about the insurer’s knowledge of individuals’ risk types. Feel free to use graphical illustration or mathematical exposition to help answer the question, but be sure to explain the intuition of your answer in words.</p> <p>B. Suppose an insurer adopted a managed care technique like prior authorization to control utilization. How would that sort of technique get incorporated into the Rothchild-Stiglitz framework, and why?</p> <p>C. A regulator can take a number of actions to help reduce adverse selection. Discuss one example of such a regulatory action. Describe how that action intends to reduce adverse selection, the strengths and weakness of that action with respect to reducing adverse selection, and any empirical evidence on the effects of that sort of action.</p>	
<p>Student’s answers (Part C)</p> <p>Weaknesses:</p> <ul style="list-style-type: none"> • 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. 	<p>Output from ChatGPT</p> <p>3. Weaknesses:</p> <ul style="list-style-type: none"> o 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. o 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.

Q3. There are many market failures in health care and health insurance.
 A. Identify three market failures that lead to large costs in terms of economic inefficiency. In two or three paragraphs for each, explain the nature of the market failure. Your answer should make the case that the welfare costs of the market failure are high, using both theoretical arguments and empirical evidence.
 B. Select one of the market failures. What public policies would effectively address this market failure? As in part (a), cite existing evidence to argue that the policy would be effective at improving economic efficiency. Please note that you may not “re-use” the answer to Question 1, Part C in this answer (i.e., you must select different examples of public policies).

Student’s answers	Output from ChatGPT
<p>First, Moral hazard leads to a loss of Pareto efficiency because of following reasons:</p> <ol style="list-style-type: none"> 1. Overconsumption of Healthcare Services: Insured individuals may consume more healthcare services than necessary because the insurance covers a significant portion of the cost. This overconsumption leads to an allocation of resources that is not optimal, as the marginal benefit of the additional healthcare services consumed is less than the marginal cost of providing them. 2. Resource Misallocation: The increased demand for healthcare services due to moral hazard can lead to a misallocation of resources within the healthcare system. Resources that could be used more efficiently elsewhere are instead used to provide services of marginal benefit to the insured individuals. 	<ol style="list-style-type: none"> 2. Moral Hazard: <ul style="list-style-type: none"> o Nature of the Market Failure: Moral hazard occurs when individuals change their behavior after obtaining insurance, becoming more likely to consume healthcare services because they do not bear the full cost of their actions. This overutilization of healthcare resources leads to economic inefficiency, as resources are diverted to low-value or unnecessary care, driving up overall healthcare costs. Moral hazard is particularly pronounced in health insurance markets because of the complex nature of healthcare decision-making and the difficulty in assessing the value of services provided. o Welfare Costs: The welfare costs associated with moral hazard are substantial. The overconsumption of healthcare services not only increases costs but also leads to the misallocation of resources, reducing the overall efficiency of the <p>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.</p>

Additional comment on this answer:
 There are a number of possible market failures that might have been used as examples in this answer. Other possibilities include principal-agent problems between doctors and patients, suboptimal consumer choices, information problems around health care prices, presence of externalities, and non-contractible quality for health care providers. All four of the faculty who reviewed the exam found it striking that the three examples selected by the student were the exact three examples that ChatGPT produced.

Q 4. “Global capitation” has emerged as a model of health care payment in recent years, where primary care organizations (for example, Oak Street Health and ChenMed) receive prospective risk-adjusted per-patient monthly payments, primarily from private Medicare Advantage plans. In return, the primary care organizations are financially responsible for **all** health care used by patients choosing to receive care from the organization (including hospital or emergency department care).
 a. How do you predict that global capitation would affect the quantity of services provided to patients by primary care physicians relative to fee-for-service reimbursement? Under what circumstances would the level of services in global capitation fall below what is socially optimal for patient health?
 b. Discuss two potential problems that could arise from unconstrained competition between primary care organizations under global capitation. What are regulations that could prevent these problems?
 c. A hypothetical empirical analysis compared health service use among Medicare beneficiaries enrolled in global capitation compared to Medicare beneficiaries whose primary care providers received fee-for-service payment, controlling for observed patient health status. The analysis found

lower overall health service use for global capitation patients and the authors concluded that global capitation reduced health care costs. Do you agree with this conclusion?

Student's answers	Output from ChatGPT
<p>Answer to part A: Circumstances Leading to Suboptimal Service Levels</p> <ol style="list-style-type: none"> 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. 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. Market Pressures: Intense competition among providers under capitation could drive some to cut corners on care quality to remain financially viable, further exacerbating the risk of under-provision. 	<p>Circumstances Leading to Suboptimal Service Levels: The level of services under global capitation could fall below what is socially optimal for patient health if:</p> <hr/> <ol style="list-style-type: none"> 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. 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. 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.
<p>Answer to part B:</p> <p>1.1 Risk Selection (Cream Skimming):</p> <p>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.</p> <p>1.4 Under-Provision of Care:</p> <p>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.</p> <p>Implementation:</p> <ul style="list-style-type: none"> 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. 	<ol style="list-style-type: none"> Risk Selection (Cream Skimming): <ul style="list-style-type: none"> 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. Under-provision of Care: <ul style="list-style-type: none"> 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.

Answer to part C:

3. **Underutilization of Necessary Services:** Lower health service use might indicate underutilization of necessary services, which could lead to poorer health outcomes and higher long-term costs. If patients are not receiving appropriate preventive care, chronic

1. **Cost Reduction vs. Underutilization:** The observed reduction in service use could indicate that global capitation encourages more efficient care, reducing unnecessary tests, procedures, and hospitalizations. However, it is also possible that the reduction in service use reflects underutilization, where necessary care is being withheld to save costs. This would be particularly concerning if the reduction in services disproportionately affects vulnerable populations or leads to worse health outcomes in the long term.
2. **Quality of Care Consideration:** Lower service use does not necessarily equate to improved efficiency if it compromises the quality of care. Without accompanying data on health outcomes, patient satisfaction, or long-term health costs, it is difficult to fully endorse the conclusion that global capitation reduces costs in a way that is beneficial for patients. For instance, if reduced service use leads to higher rates of hospitalization or more severe health issues in the future, the short-term cost savings could be offset by long-term health costs.

disease management, or timely interventions, their conditions might worsen, leading to more severe and costly health episodes in the future.

4. **Adverse Health Outcomes:** Reductions in health service use without corresponding improvements in health outcomes could imply that patients are foregoing essential care. This could result in increased morbidity and mortality, which are not immediately reflected in short-term cost savings but may lead to higher costs over time.

Additional comments on this answer:

1. In Part A, the student gave, "Market pressure" as part of their answer. Whereas the other instances of circumstances that were provided seem to relate closely to the ChatGPT output, the answer around "Market pressure" is distinctly incorrect.
2. As shown in the screenshot above from Part B, the student used the acronym "PCO" throughout their answer. This acronym was produced by ChatGPT. However, it is not a standard acronym that is used in our field at all.
3. Conducting audits and patient satisfaction surveys was produced by ChatGPT in its answer to Part C, and appear in the student's answer, as well. These are not things that we think of as a standard answer to the question, particularly from the lens of Health Economics. To our knowledge, these sorts of things were not covered in our classes, nor in the readings. While the answer may be somewhat relevant, listing "Regular Audits" and "Patient Satisfaction Surveys" as the first two items under "Implementation" strikes us very unusual in the context of this exam.

4. As the AOE faculty continued to assess the prelim exam answers, we wanted to compare the prelim answers with other writing that the student had submitted.

As background, students are always given prior health economics prelim exams to help them prepare for their own exam. This student received those prior exams, following standard practice. However, this student wanted to share their answers to prior prelim questions and receive specific feedback from several of the health economics AOE faculty. This is not standard practice. Nevertheless, several faculty members did meet with the student individually and gave feedback on the "practice" questions, although it was eventually communicated by some AOE faculty that this was not standard and that the pace of this type of feedback would need to be limited.

One AOE faculty had a copy of the student's "practice" answers to the 2017 prelim exam that were sent via email. Upon reviewing those answers, the faculty member noted similarities with the answers on the actual prelim. When the faculty member put the 2017 prelim questions into

ChatGPT, there was strong evidence that the student used ChatGPT as the basis for their answers to the “practice” prelim questions. Specific examples:

<p>Question 4: Medicare programs, many state Medicaid programs and the Affordable Care Act Marketplaces use the combination of capitation and risk adjustment to pay insurers.</p> <p>a. Describe pure capitation, and briefly state the incentives this payment mechanism creates on health spending</p> <p>b. Describe risk adjustment, and discuss what problem(s) it intends to address. Also discuss how insurers may potentially respond to risk adjustment.</p>	
<p>Student’s answer</p>	<p>ChatGPT’s answer</p>
<p>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.</p>	<div style="border: 1px solid gray; padding: 10px; margin-bottom: 10px;"> <p>Question 4. Medicare program, many state Medicaid programs and the Affordable Care Act Marketplaces use the combination of capitation and risk adjustment to pay insurers.</p> <p>a. Describe pure capitation, and briefly state the incentives this payment mechanism creates on health spending</p> <p>b. Describe risk adjustment, and discuss what problem(s) it intends to address. Also discuss how insurers may potentially respond to risk adjustment.</p> </div> <p>Certainly! Let's break down both concepts:</p> <p>a. Pure Capitation</p> <p>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.</p>
<p>Incentives Created:</p> <ul style="list-style-type: none"> ● 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 	<p>Incentives Created by Pure Capitation:</p> <ol style="list-style-type: none"> 1. 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. 2. 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. 3. 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. 4. 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.

<p>unnecessary services since providers bear the financial risk.</p> <ul style="list-style-type: none"> ● Risk Selection: May lead to risk selection, where providers might avoid enrolling individuals who are expected to have high healthcare costs. 	
<p>Risk Adjustment</p> <p>Definition: Risk adjustment is a mechanism used to compensate health insurance plans for the differences in risk profiles of their enrollees. The goal is to adjust payments to health plans based on the health status and expected costs of the enrollees they cover.</p>	<p>b. Risk Adjustment</p> <p>Description: Risk adjustment is a method used to account for the health status and risk profile of individuals when determining payments to insurers or providers. The goal is to adjust payments based on the expected cost of providing care to individuals with varying levels of health risk. For instance, individuals with chronic conditions or complex health needs typically require more resources, so payments are adjusted to reflect these higher costs.</p>

5. The AOE faculty are in possession of a final exam, which was handwritten by the student in question for PUBH 6832 in the fall of 2022. We include this here as an attachment, as a known example of writing in the student’s voice. We observe many dissimilarities between this example of the student’s writing and the submission for the preliminary exam.
6. Overall, the health economics AOE faculty agree that there is a preponderance of evidence that LLM-tools were used in the student’s prelim exam, despite being explicitly prohibited. This conclusion is based on concerns of directly using such a LLM tool for the specific prelim questions that were asked, as well as indirectly via the “practice” prelim questions that they received feedback on from AOE faculty.

Attachments:

- a. The student's submitted preliminary exam.
- b. The preliminary exam that was given to the student.
- c. The output of the prelim exam answers, as produced by ChatGPT.
- d. The "practice prelim" 2017 exam that the student sent to one of the AOE faculty for feedback.
- e. A copy of the student's prior written exam from another course, where it is known that no LLMs were used in the answer.