Using AI as Gatekeeper or Second Opinion: Designing Patient Pathways for AI-Augmented Healthcare*

Tinglong Dai† Simrita Singh‡

†Carey Business School, Johns Hopkins University, Baltimore, Maryland 21202;
Hopkins Business of Health Initiative, Johns Hopkins University, Washington, District of Columbia 20001, dai@jhu.edu
‡Leavey School of Business, Santa Clara University, Santa Clara, California 95053, ssingh17@scu.edu

Of the 950 artificial intelligence (AI) systems cleared by the U.S. Food and Drug Administration as of June 2024, most function as classifiers to help screen or diagnose specific medical conditions. Yet, questions remain about how to best integrate AI into healthcare workflows, including whether AI should serve as a gatekeeper, determining which patients require human attention, or as a second opinion to complement medical consultations. Motivated by this question, we model a healthcare system in which patients can consult a specialist, an AI system, or both. The key design question is whether the patient should first consult AI or the specialist, corresponding to AI's gatekeeper and second-opinion roles, respectively. We model a two-step decision-making process influenced by an initial signal, or anchor. Contrary to popular belief, we show using AI as a gatekeeper does not necessarily increase missed diagnoses; using AI as a second opinion, on the other hand, reduces missed diagnoses but can also increase false positives. In general, the gatekeeper approach is preferable in low-risk settings, whereas the second-opinion approach is better suited for high-risk patients for whom avoiding missed diagnoses is a primary concern. Notably, scenarios exists where AI should *not* be used for intermediate-risk patients for whom uncertainty is highest, challenging the premise that AI is most useful in reducing uncertainty. Finally, applying our model to glaucoma diagnosis, we numerically illustrate cost savings from optimizing patient pathways. Our work highlights the potential for AI to contribute to the United Nations' Sustainable Development Goals by optimizing resource allocation and improving patient outcomes.

Key words: Artificial intelligence, healthcare systems, service design, human-AI interaction.
History: This version is dated December 14, 2024

1. Introduction

The clock is ticking for the United Nations' Sustainable Development Goals (SDGs), a set of 17 goals aimed at improving economic, environmental, and social conditions worldwide by 2030 (Sodhi and Tang 2024). The third of these goals is to "ensure healthy lives and promote well-being for all

* Tinglong Dai and Simrita Singh contributed equally to this paper and are listed in alphabetical order.

at all ages." Achieving this goal requires innovative solutions to bridge gaps in healthcare delivery, particularly in low-resource settings (Chen et al. 2021, Lee et al. 2013, Lee and Tang 2018, Mehta et al. 2016). Among these innovations, artificial intelligence (AI) stands out as a transformative approach (Abràmoff et al. 2024, Dai and Tayur 2022). A compelling example of AI's potential is its application in screening for diabetic retinopathy (DR), a serious complication of diabetes that can lead to vision loss. Approximately 11% of the U.S. population has diabetes (CDC 2024), which is similar to the global prevalence (Sun et al. 2022). DR is a serious complication of diabetes that can result in vision loss. Yet, less than 15% of diabetic patients receive the recommended annual screening, primarily due to the cost and inconvenience of scheduling appointments with the few available eye care specialists (Dai and Abràmoff 2023). AI-based devices, such as IDx-DR, with their low variable costs and high diagnostic accuracy, can autonomously diagnose DR and flag positive cases for human specialists.

Incorporating AI into routine healthcare delivery systems, however, requires purposeful design (Dai and Tayur 2022). Of the 950 AI systems cleared by the U.S. Food and Drug Administration (FDA) as of June 2024 (FDA 2024), most function as classifiers to help diagnose specific medical conditions. However, the role of medical AI in healthcare delivery remains a subject of considerable debate. In the U.S., many AI devices such as IDx-DR have been used primarily as *gatekeepers* at the point of care (Dai and Abràmoff 2023). For example, patients who receive negative screening results from IDx-DR are not referred to an ECP for a definitive diagnosis. An alternative approach is to refer all patients to the ECP for diagnosis. Many ECPs do not have IDx-DR in their offices, so they will make diagnoses without recourse to AI. If an ECP does have access to IDx-DR, they can use IDx-DR after their encounter with the patient. In this "specialist-first" approach, AI serves as a *second opinion*.

One potential concern about using AI as a gatekeeper is the risk of false negatives, where a health condition of concern is missed (Dai and Tayur 2022). However, positive screening results can provide additional information to healthcare professionals to help them make informed decisions and confirm diagnoses. When a screening test returns a positive result, it raises awareness among providers, leading to a reassessment that ensures accurate diagnoses and appropriate care for patients.

Compared with gatekeeping, AI is widely believed to best serve as a second opinion. For example, Ho et al.'s 2022 survey of 252 Australian eye care professionals shows a strong preference for using AI as an additional tool *after* consultation (i.e., as a second opinion) rather than for an initial diagnosis (i.e., as a gatekeeper), in part because of concerns that inaccurate AI advice may lead to false-negative diagnoses: A patient with a DR condition may receive a false-negative screening result, preventing the patient from being referred to the ECP. Similarly, a patient who does not have DR may receive a false-positive screening result from the AI system, potentially leading to unnecessary tests and procedures.

Our work is motivated by the ongoing debate on whether AI should be used as a gatekeeper or as a second opinion. Specifically, we seek to answer the following questions: (1) Compared with the case without AI, what are the diagnostic performance and patient outcome implications of using the "AI first" approach? (2) Compared with the case without AI, what are the diagnostic performance and patient outcome implications of using the "specialist first" approach? (3) What's the optimal design choice between the "AI first" and "specialist first" approaches? To answer these questions, we model and analyze the diagnostic performance and impact on patient outcomes under these two approaches, generating insights into when AI should be used as a gatekeeper or as a second opinion. For both approaches, inspired by the literature on judgment and decision-making (e.g., Hastings 1970, Ogdie et al. 2012, Tversky and Kahneman 1974), our model incorporates the decision process that depends on an initial signal, the so-called anchoring effect. This effect is relevant because the order in which the specialist and AI see the patient affects the specialist's initial impression. Specifically, the first diagnosis generated tends to become an "anchor" that can bias later assessments. Thus, if AI screens first, the specialist may anchor to that initial result, even if it is inaccurate. Conversely, if the specialist examines the patient first, his initial diagnosis may anchor AI's algorithmic assessment.

First, we show that, contrary to popular belief (Ho et al. 2022), using AI as a gatekeeper does not necessarily increase missed diagnoses compared with specialist-first diagnoses. The reason is that when AI is used as a gatekeeper, in some cases, a specialist with an inaccurate diagnosis may benefit from having an anchor from AI signal that is accurate. In particular, when AI has a higher sensitivity than the specialist, using AI as a gatekeeper can help reduce the chance of false-negative diagnoses. For the same reason, using AI as a gatekeeper does not necessarily reduce false-positive diagnoses and thus unnecessary treatments. The reason is that AI could lead the specialist to overdiagnose a condition that the patient does not have.

Second, we show using AI as a second opinion leads to fewer false-negative diagnoses than using AI as a gatekeeper or not using AI. By having the specialist examine the patient first, the anchoring to a false-negative diagnosis is mitigated. The specialist can then use AI as a second opinion to fill in gaps and reduce human oversight. However, using AI as a second opinion may or may not reduce false-positive diagnoses compared with not using AI. For example, suppose the prior confidence is high, but the specialist makes a true negative diagnosis and AI makes a false-positive diagnosis. In this case, AI's contradictory diagnosis, given the high prior, can cause the specialist to accept AI's diagnosis.

Third, whether AI should be used as a gatekeeper or as a second opinion depends on whether avoiding false negatives or false positives is more important in a given clinical context. Taken together, we find using AI as a gatekeeper is most beneficial for low-risk patients, for whom the downside of false negatives is lower. Using AI as a second opinion, on the other hand, is optimal for high-risk patients, for whom missing a diagnosis has serious consequences. Although this finding may seem intuitive at first glance, the implication is surprising: These results suggest AI should not be used when patients are neither high nor low risk. Put differently, AI should be avoided altogether, neither as a gatekeeper nor as a second opinion, for the most uncertain patient groups. This finding contradicts the intuition that, in the absence of agency problems, the role of AI is to provide information to supplement physician decision-making, and that such information is most valuable for intermediate-risk patients, for whom uncertainty is highest (Dai and Tayur 2022).

The decision pattern of avoiding AI for the most uncertain patient groups has been reported in the literature, albeit driven by different underlying mechanisms. Price et al. (2019) argue physicians may bear additional legal liability for disagreeing with AI recommendations. Because they are most likely to disagree with AI for the most uncertain patient populations, they avoid using AI in these cases to mitigate potential legal liability for using (but not following) AI. Dai and Singh (2020) provide another mechanism that explains the same phenomenon; that is, physicians avoid using AI when it is most needed to signal their inherent diagnostic abilities. In our model, physicians do not face legal liability, nor do they have an incentive to signal their skills. Rather, the anchoring effect is what reduces the informational benefit of obtaining additional information from AI to the extent that it is optimal not to use it for the most uncertain patient populations. This effect provides a novel mechanism for an important concern about the underuse of AI in medical decision making.

Our theoretical result, rooted in the psychological concept of anchoring, suggests physician-AI collaboration may, in some instances, yield worse outcomes than physician decision-making alone. Empirical evidence supports this finding: in a recent randomized clinical trial of 50 physicians (Goh et al. 2024), the use of an AI chatbot failed to significantly improve diagnostic performance, largely because physicians frequently dismissed the chatbot's recommendations when they contradicted their own initial diagnoses.

Finally, we analyze a real-world dataset of glaucoma patients to evaluate the benefits of incorporating prior probabilities and the anchoring effect in selecting optimal patient-specific pathways. We use a probit model to predict the likelihood of glaucoma based on demographic and clinical characteristics and compare the total expected costs of a missed diagnosis and unnecessary treatment under three standard pathways: no AI, AI as a gatekeeper, or AI as a second opinion for all patients. For low specialist sensitivity and high cost of a missed diagnosis, using a highly sensitive AI as a second opinion can reduce costs by approximately 38%-68% relative to no AI and by

23%-46% relative to AI as a gatekeeper. When the cost of a missed diagnosis equals unnecessary treatment, using AI as a gatekeeper reduces costs by about 25%-45% relative to no AI and by 23%-45% relative to AI as a second opinion. For specialists with high sensitivity, an indiscriminate AI use can worsen performance, especially when the cost of a missed diagnosis is significantly higher than the cost of unnecessary treatment.

Our research contributes to the growing effort to improve access to healthcare services in resource-constrained settings by operationalizing AI-enhanced patient pathways. Similar to the approaches explored by Olsder et al. (2023) in improving the availability of treatments for rare diseases through outcome-based payment schemes, and by Lee et al. (2013), Mehta et al. (2016), and Chen et al. (2021), among others, in facilitating large-scale healthcare delivery in rural areas through vehicle management, our work aims to reduce the cost of access and improve the quality of healthcare. By deploying AI as a gatekeeper or second opinion, we address critical gaps in healthcare efficiency and provide practical pathways for integrating AI into routine medical decision-making, particularly where healthcare resources are limited.

Our work also advances the understanding of the optimal sequencing of human and AI interventions in service operations, drawing a parallel to optimal design problems in supply chain and service operations. Inspired by Lee and Tang (1997), who explore the intertemporal trade-offs of product differentiation, we consider expedited versus delayed patient differentiation: using AI as a gatekeeper leads to earlier patient classification, whereas using AI as a second opinion defers this decision. We investigate the impact of different AI integration points on healthcare outcomes and provide insights into how AI can be effectively embedded in healthcare workflows to improved patient outcomes.

2. Literature

Our paper builds on and advances several streams of literature: sequencing and cognitive biases, quantitative models of sequential decision-making, human-AI interaction, and service design. In addition, our paper relates to the medical literature on the role of AI in clinical decision-making.

First, anchoring bias is among the most robust cognitive effects documented in psychology (Tversky and Kahneman 1974). This bias occurs when decision-makers rely disproportionately on the first piece of information (the "anchor") when forming judgments, even when subsequent information is available. Anchoring bias has been observed across diverse settings involving quantitative judgments and is recognized as a key cognitive bias contributing to diagnostic errors in medical decision-making (Ogdie et al. 2012). Kahneman and Miller (1986) introduce a norm theory in which events are evaluated against a reference norm—a consolidation of past experiences. Events resembling the norm are perceived as unsurprising, with their evaluation anchored to this baseline. Conversely, events that deviate markedly from the norm elicit surprise.

Anchoring bias is widely acknowledged in the medical community. Ly et al. (2023) provide empirical evidence that emergency room physicians are influenced by the patient's stated reason for visit—information recorded prior to the physician's evaluation—resulting in underdiagnosis of conditions not mentioned in the initial presentation. Rastogi et al. (2022) explore the effects of anchoring bias on diagnostic accuracy through experimental studies. Similarly, Bach et al. (2023) qualitatively examine how anchoring bias manifests in clinical settings with the introduction of AI support systems. To the best of our knowledge, this study is the first to investigate the impact of anchoring bias on diagnostic decisions that involve both a specialist and a diagnostic tool.

Second, the literature has established a mathematical foundation for modeling sequential decision-making. Jin et al. (2023) extend norm theory (Kahneman and Miller 1986) by modeling the norm as a weighted average of the risk levels of previous patients. In their framework, physicians adjust the risk assessment of a new patient relative to this norm. The adjustment varies with the perceived difference between the anchor (norm) and the current case. Small differences result in estimates positively correlated with the anchor, whereas large differences lead to overcompensation, producing negative correlations. Jin et al. (2023) validate their model using a large dataset, providing empirical evidence of anchoring bias in the form of positive auto-correlation in physician decisions. For instance, physicians were more likely to admit patients or order additional tests for a current patient if they had done so for the previous one.

Lieder et al. (2012) develop and validate a Metropolis-Hastings model to describe the psychological process underlying probabilistic inference as a sequence of adjustments to an initial anchor (anchoring-and-adjustment). In this model, adjustments that increase the estimate's likelihood are always accepted, while less probable adjustments are accepted with a probability proportional to their posterior likelihood relative to the anchor.

Although Jin et al. (2023) and Lieder et al. (2012) propose distinct approaches—a risk/utility-based model and a probabilistic choice/acceptance model, respectively—their core principle is similar: adjustments significantly deviating from the anchor override it entirely, whereas smaller deviations retain a strong residual influence of the anchor. Building on the Metropolis-Hastings algorithm, our paper develops a tractable framework for modeling sequential diagnostic decision-making.

Third, and most relevant to our work, the literature has examined human-AI interaction and related service-design principles; see Dai and Tayur (2022) for an overview of research opportunities in this area. Studies typically focus on designing AI systems that account for human behavior (Grand-Clèment and Pauphilet 2023), analyzing the interactive decision-making process (de Véricourt and

Gurkan 2023), or triaging patient cases to determine whether AI or human physicians should handle them (Dvijotham et al. 2023). For instance, Orfanoudaki et al. (2022) use surveys to identify optimal designs for human-AI collaboration. However, empirical evidence highlights the challenges of such collaboration. In a randomized clinical trial involving 50 physicians (Goh et al. 2024), AI chatbot recommendations did not consistently improve diagnostic accuracy, because physicians often disregarded suggestions that conflicted with their initial diagnoses.

Several recent papers shed new light on AI-augmented healthcare Adida and Dai (2024) investigate how different physician payment systems influence the use of confirmatory tests (or AI) in healthcare diagnostics. Fügener et al. (2022) find human-AI collaboration is most effective when AI delegates tasks to humans, but not vice versa. Similarly, Agarwal et al. (2024) study human-AI collaboration in radiology, showing AI's confident predictions enhance radiologists' performance, yet its performance declines when AI provides uncertain recommendations. Lai et al. (2024) explore regulatory approaches for medical AI, emphasizing the strategic role of flexible oversight in encouraging compliance during algorithm retraining. Luan et al. (2024) analyze liability frameworks addressing AI biases and find stringent liability can unintentionally lead to biased usage patterns among healthcare providers. Mullainathan and Obermeyer (2021) highlight how systematic errors in risk estimation can lead to both under-testing and over-testing, emphasizing the importance of incorporating physician errors into decision models—a perspective consistent with our modeling of anchoring bias. Dai and Singh (2024) examine how physicians' liability concerns shape their decisions to use AI in treatment planning, showing current liability frameworks can lead to both overuse and underuse of AI. Whereas our research aligns with these studies, our focus differs: we examine how anchoring bias affects the optimal sequencing of AI and specialists in diagnostic workflows.

Our work also relates to the design and analysis of gatekeepers in service systems. The analytical literature in this domain (see, e.g., Shumsky and Pinker 2003) focuses on the gatekeeper's optimal transfer (to the service expert) response to different incentive schemes and congestion levels. Hathaway et al. (2023) empirically test the predictions (hypotheses) of such an analytical model. Freeman et al. (2021) use data from visits to an emergency department (ED) to show a gatekeeping unit (that decides patient discharge or admission to the hospital) reduces both unnecessary hospitalizations and wrongful patient discharges. To our best knowledge, we are the first to analytically model the anchoring effect of using a gatekeeper and its implication for the quality of diagnosis (overall missed diagnosis and unnecessary treatment). Further, we investigate the settings and patient risk levels under which using a gatekeeper is beneficial.

Beyond healthcare, Balakrishnan et al. (2022) conduct laboratory experiments to investigate the factors driving over- and under-adherence to algorithmic recommendations when humans possess private information. Cui et al. (2022) use field experiments on online platforms to compare supplier pricing behavior when responding to human buyers versus AI-based chatbots. Our work is distinct in both methodology and research focus, as we analyze and optimize the role of AI in patient pathways, accounting for the cognitive impact of anchoring bias. Gurkan and de Véricourt (2022) examine contracting and pricing decisions related to AI adoption, while Miao et al. (2024) provide empirical evidence on how AI investment in human capital positively affect focal firms and generate spillover benefits for their suppliers.

Our work also relates to the service operations literature on the optimal sequencing of activities to maximize the total service utility of the customer under acclimation, memory decay, and higher memorability of the peak event (Das Gupta et al. 2016, Li et al. 2022). Our work is distinct in that we do not optimize for customer experience but rather for the costs of diagnostic errors under service sequencing when the first outcome serves as an anchor for the final diagnosis.

Finally, our paper contributes to the medical AI literature, which increasingly focuses on scaling AI in healthcare delivery (Abràmoff et al. 2024). Recent clinical trials and observational studies have explored the real-world performance of AI in medical practice. For example, Mathenge et al. (2022), in a randomized controlled trial in Rwanda, find that AI-supported DR screening with immediate results increased referral adherence by 30% in the intervention group compared to the control group. Ruamviboonsuk et al. (2019) report that a deep learning algorithm achieved significantly higher sensitivity (0.97 vs. 0.74) and slightly lower specificity (0.96 vs. 0.98) than human graders for DR screening. Xie et al. (2020) conduct a data-driven economic analysis demonstrating semi-automated AI screening for DR is the most cost-effective compared to fully automated AI or physician-only screening. Similarly, Wolf et al. (2020) and Ahmed et al. (2024) demonstrate autonomous AI systems for DR screening can be both cost-saving and cost-effective for children with diabetes at the individual patient and system levels. In a more recent randomized controlled trial in Bangladesh, Abràmoff et al. (2023) show using an autonomous AI system as a triaging tool at an eye hospital increases physician productivity by 40%. Relevant to our focus on AI as a second opinion, American College of Radiology (2021) present a case study on the impact of AI in diagnosing and triaging radiology images at a large medical center, highlighting improvements in workflows and radiologists' learning. These findings motivate our exploration of how AI-generated diagnoses can augment physicians' diagnostic decision-making.

3. Model

We study sequential decision-making by AI and the specialist, where the final decision-maker, the specialist, assimilates two independent pieces of information: the assessment of AI and the specialist's own independent assessment to make the final judgment.

3.1. Anchoring and Adjustment

We use a two-step anchoring and adjustment model, as presented in Lieder et al. (2012), to capture the diagnostic decision-making under sequential and independent assessments by AI and a specialist. For example, when AI is used as a gatekeeper, the patient is referred to a specialist only if AI generates a positive screening result; in this scenario, AI's diagnosis acts as an anchor and the specialist's diagnosis acts as an adjustment to the anchor. Conversely, when AI is used as a second opinion, the specialist's diagnosis is the anchor and AI's diagnosis is the adjustment. If the strength of the adjustment (the probability that the true condition aligns with the adjustment given the anchor) is greater than the strength of the anchor (the probability that the true condition aligns with the anchor given the adjustment), the adjustment is always accepted. Otherwise, the acceptance probability of the adjustment is equal to the ratio of the strength of the adjustment to the strength of the anchor.

We consider a two-step process whereby an initial diagnosis (the anchor) is followed by another independent diagnosis (the adjustment). The acceptance probability of the adjustment x' over the anchor x is defined as

$$A(x'|x) = \begin{cases} 1 & if\ x' = x \\ \min\left\{1, \frac{\Pr(x')\Pr(x|x')}{\Pr(x)\Pr(x'|x)}\right\} & if\ x' \neq x \end{cases}. \tag{1}$$

Here, $\Pr(y)$ represents the probability of the true condition being y, and $\Pr(z|y)$ represents the conditional probability of observing z given the true condition y. We refer to $\Pr(x')\Pr(x|x')$ as the strength of the adjustment and $\Pr(x)\Pr(x'|x)$ as the strength of the anchor. The model anchor influences the final decision as follows: when the adjustment equals the anchor, the anchor is retained. However, when the anchor differs from the adjustment, it still overrides the adjustment with a residual probability (equal to 1-acceptance probability).

Notably, (1) coincides with the Metropolis-Hastings algorithm (Hastings 1970, Lieder et al. 2012) that has been used in the literature to capture the psychological process underlying probabilistic inference as a sequence of adjustments to an initial guess. In our setting, the potential states are denoted as $\{1,0\}$, representing positive and negative states, respectively. Let α and β denote the sensitivity (probability of positive diagnosis when the true state is 1) and the specificity (probability of negative diagnosis when the true state is 0), respectively.

Because the AI system and the specialist differ in their sensitivity ($\alpha_{AI},\ \alpha_S$) and specificity ($\beta_{AI},\ \beta_S$), the probability $P(z|y)$ of a diagnosis z given the true state y depends on whether the diagnosis is made by the AI or the specialist. Table 1 presents the probabilities of the two true states ($y = 0, 1$) alongside the probabilities of positive and negative diagnoses by the AI and the specialist for each state.

True State (z)	Diagnosis (y)	$P(z)$	$P(y\|z)$ with AI	$P(y\|z)$ with Specialist
1	0	p	$1-\alpha_{\text{AI}}$	$1-\alpha_S$
1	1	p	α_{AI}	α_S
0	0	$1-p$	β_{AI}	β_S
0	1	$1-p$	$1-\beta_{\text{AI}}$	$1-\beta_S$

Table 1 Probabilities of true states and diagnoses by AI and the specialist.

Example 1. Suppose AI first generates a positive diagnosis, and then the specialist generates a negative diagnosis. The acceptance probability of the specialist's negative diagnosis is using (1),

$$A(S=0|AI=1)=\min\left\{1,\frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right\}.$$

The acceptance probability is higher for smaller values of p, indicating a lower prior probability that the patient has the condition. In addition, a higher value of α_S, implying the specialist rarely misdiagnoses positive cases, increases the acceptance probability. Conversely, a higher β_{AI} value, implying AI rarely gives false-positive diagnoses, decreases the acceptance probability. Similarly, we can obtain the acceptance probabilities $A(S=1|AI=0)$, $A(AI=0|S=1)$, and $A(AI=1|S=0)$ (see the appendix).

3.2. How Anchoring and Adjustment Make Sequencing Matter

Under the anchor-adjustment model, the sequence in which diagnoses are made by AI and/or the specialist influences the likelihood of receiving a positive or negative diagnosis.

Example 2. Consider a case where AI gives a positive diagnosis but the specialist gives a negative diagnosis. Under the anchoring and adjustment model, the probability of the final diagnosis depends on whether the patient saw AI or the specialist first.

- AI as a gatekeeper (AI first): The probability of retaining AI's initial positive diagnosis over a second negative diagnosis by the specialist is

$$1-\min\left\{1,\frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right\}.$$

- AI as a second opinion (specialist first): The probability of accepting a positive diagnosis by AI after an initial negative diagnosis by the specialist is

$$\min\left\{1,\frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\}.$$

In Example 2, consider a case where the independent diagnoses by the AI and the specialist are positive and negative, respectively, and $\frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}>1$, which holds when $p<\frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}$. Under these conditions, the probability of a positive diagnosis is zero when the AI acts as a gatekeeper. In contrast, when the AI provides a second opinion, the probability of a positive diagnosis is $\frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}$.

3.3. Model Limitations

Our model assumes the diagnoses made by the specialist and AI are independent of the prior. This assumption, though potentially counterintuitive in extreme cases, proves analytically useful. For example, if $p=0$ (indicating no probability of disease), the specialist could still make a positive diagnosis with probability $1-\beta_S$. Similarly, if $p=1$, the specialist could give a negative diagnosis with probability $1-\alpha_S$. However, we contend that in scenarios with extremely low or high priors, the question of diagnosis—and consequently the optimal clinical path for the patient—becomes largely irrelevant and thus beyond the scope of our paper.

A more realistic assumption would posit that the prior p ensures the specialist's diagnosis aligns with economic rationality under Bayesian posterior beliefs. Specifically, if the specialist makes a positive diagnosis, their posterior belief would justify flagging the disease as present as the "economically" optimal decision. Similarly, a negative diagnosis would correspond to a posterior belief that makes it "economically" optimal to conclude the disease is absent. Equivalently,

$$\frac{p\alpha_S}{p\alpha_S+(1-p)(1-\beta_S)}C_M \geq \frac{(1-p)(1-\beta_S)}{p\alpha_S+(1-p)(1-\beta_S)}C_T \iff p \geq \frac{(1-\beta_S)C_T}{\alpha_S C_M+(1-\beta_S)C_T},$$
$$\frac{p(1-\alpha_S)}{p(1-\alpha_S)+(1-p)\beta_S}C_M \leq \frac{(1-p)\beta_S}{p(1-\alpha_S)+(1-p)\beta_S}C_T \iff p \leq \frac{\beta_S C_T}{(1-\alpha_S)C_M+\beta_S C_T}.$$

If we impose the same restrictions on AI, we obtain

$$\begin{aligned}
&\underline{p} \leq p \leq \bar{p},\\
&\underline{p} = \max\left\{\frac{(1-\beta_{AI})C_T}{\alpha_{AI}C_M+(1-\beta_{AI})C_T}, \frac{(1-\beta_S)C_T}{\alpha_S C_M+(1-\beta_S)C_T}\right\},\\
&\bar{p} = \min\left\{\frac{\beta_{AI}C_T}{(1-\alpha_{AI})C_M+\beta_{AI}C_T}, \frac{\beta_S C_T}{(1-\alpha_S)C_M+\beta_S C_T}\right\}.
\end{aligned}$$

This regime implies both the specialist and the AI declare positive outcomes as positive and negative outcomes as negative, consistent with their Bayesian posterior beliefs. We argue imposing these restrictions on the prior p does not alter the qualitative insights derived from our analysis. Furthermore, if $\frac{1-\beta_S}{\alpha_S}$ and $\frac{1-\beta_{AI}}{\alpha_{AI}}$ are sufficiently small, and $\frac{1-\alpha_S}{\beta_S}$ and $\frac{1-\alpha_{AI}}{\beta_{AI}}$ are similarly small, the range of p satisfying $\underline{p} \leq p \leq \bar{p}$ spans nearly the entire interval $(0,1)$.

4. Optimal Patient Pathway Design with AI

The use of AI in a healthcare delivery system has been proposed in two possible configurations—as a gatekeeper (to selectively refer patients to the specialist) or as a second opinion (to confirm or refute the specialist's diagnosis). The potential benefit of the first approach is to reduce the cost of specialist consultation, especially in the presence of capacity constraints. The potential benefit of the second approach is improved accuracy. Despite the potential benefits of AI, the perceived risks of missed diagnoses and unnecessary treatment due to AI act as barriers to the adoption of AI. Section 4.1 and Section 4.2 analyze these risks when using AI as a gatekeeper and as a second opinion, respectively. Section 4.3 compares both strategies (and the case without AI) to derive the condition for each strategy to be the least risky. For the remainder of the paper, we assume both AI and the specialist are imperfect ($\alpha_S,\ \beta_S,\ \alpha_{AI},\ \beta_{AI} < 1$).

4.1. AI as a Gatekeeper

A key concern with using AI as a gatekeeper is its potential to restrict certain patients' access to specialists, thereby increasing missed diagnoses. At the same time, AI as a gatekeeper could reduce unnecessary consultations and treatments. Proposition 1 demonstrates that whereas AI as a gatekeeper may increase the probability of missed diagnoses in certain scenarios, it can, counterintuitively, reduce the probability of missed diagnoses in others.

Proposition 1 *If $\alpha_{AI} \le \alpha_S$, then for all priors p, AI as a gatekeeper has a higher probability of a false-negative diagnosis (missed diagnosis) than the specialist alone. Otherwise, if $\alpha_{AI} > \alpha_S$, a threshold $\tau_M^G \in (0,1]$ exists:*

$$\tau_M^G = \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\frac{\alpha_S}{\alpha_{AI}})}.$$

1. *AI as a gatekeeper reduces the overall probability of a false-negative diagnosis (missed diagnosis) if and only if $p \in (\tau_M^G, 1]$.*
2. *AI as a gatekeeper increases the overall probability of a missed diagnosis for all priors $p \in [0, \tau_M^G)$.*
3. *The threshold τ_M^G is increasing in α_S but decreasing in α_{AI} and β_{AI}.*

For a given prior p, the probability of a missed diagnosis when the patient is seen only by the specialist, denoted $P_M^0(p)$, is the probability that the patient has the disease and the specialist provides a negative diagnosis; that is,

$$P_M^0(p) = p(1-\alpha_S). \tag{2}$$

In a gatekeeper AI system, a missed diagnosis can occur in two ways: (1) AI misdiagnoses a positive case as negative, or (2) AI correctly identifies a positive case, but the specialist subsequently makes a false-negative diagnosis and overrides AI's positive diagnosis. The probability of a missed diagnosis with AI as a gatekeeper is

$$P_M^G(p) = p(1-\alpha_{AI}) + p\alpha_{AI}(1-\alpha_S)\min\left\{1, \frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right\},$$

which simplifies to:

$$P_M^G(p) = \begin{cases} p(1-\alpha_{AI}\alpha_S) & \text{if } p \leq \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}, \\ p(1-\alpha_{AI}) + (1-p)\alpha_{AI}(1-\beta_{AI}) & \text{otherwise.} \end{cases} \tag{3}$$

The use of AI as a gatekeeper introduces two competing effects. First, it can increase the number of false negatives (missed diagnoses) beyond those made by the specialist alone. Second, it can reduce missed diagnoses by correcting some of the specialist's errors, as AI's positive diagnosis serves as an anchor that influences the specialist's decision-making. The dominance of these effects depends on AI's sensitivity (α_{AI}) and the prior (p). If $\alpha_{AI} \leq \alpha_S$, AI will not reduce missed diagnoses, because the first effect dominates. When $\alpha_{AI} > \alpha_S$ and the prior p is small, the specialist is more likely to disregard AI's positive diagnosis (see Example 2). However, as p increases, a positive diagnosis from AI becomes a stronger anchor, making the specialist less likely to uphold a negative diagnosis. In this case, the second effect dominates, leading to fewer missed diagnoses.

The strength of this anchoring effect relative to the adjustment depends on AI's specificity and the specialist's sensitivity. The viable region where gatekeeper AI can reduce missed diagnoses expands with higher AI specificity (fewer false-positives) and decreases with the specialist's sensitivity (fewer false negatives). Additionally, the viable region grows as AI's sensitivity α_{AI} increases, because higher sensitivity lowers the probability of AI incorrectly screening out positive cases.

Next, we show that, contrary to expectation, AI as the gatekeeper may increase the overall probability of a false-positive diagnosis and may thus lead to unnecessary treatment.

Proposition 2 *If $\beta_{AI} \geq \beta_S$, AI as a gatekeeper decreases or maintains the probability of a false-positive diagnosis (unnecessary treatment) across all priors p. Conversely, if $\beta_{AI} < \beta_S$, a threshold $\tau_T^G \in (0,1]$ exists:*

$$\tau_T^G = \frac{(1-\beta_{AI})^2}{(1-\beta_{AI})^2 + (1-\alpha_S)(1-\frac{\beta_{AI}}{\beta_S})}.$$

1. *Using AI as a gatekeeper increases the probability of a false-positive diagnosis for all priors $p \in (\tau_T^G, 1]$.*
2. *Using AI as a gatekeeper reduces the probability of a false-positive diagnosis for all priors $p \in [0, \tau_T^G]$.*
3. *The threshold τ_T^G increases with α_S but decreases with β_S. It decreases with β_{AI} if $\beta_{AI} < 2\beta_S - 1$, and increases with β_{AI} otherwise.*

For a given prior p, the probability of unnecessary treatment when the patient is seen only by the specialist, denoted $P_T^0(p)$, is the probability that the patient does not have the disease and the specialist provides a positive diagnosis

$$P_T^0(p) = (1-p)(1-\beta_S). \tag{4}$$

Under a gatekeeper AI system, unnecessary treatment can occur in two scenarios: (1) when both AI and the specialist incorrectly diagnose a negative case as positive, or (2) when AI incorrectly diagnoses a negative case as positive, and the specialist correctly diagnoses it as negative but retains AI's initial positive diagnosis. The overall probability of unnecessary treatment with AI as a gatekeeper is:

$$P_T^G(p) = (1-p)(1-\beta_{AI})(1-\beta_S) + (1-p)(1-\beta_{AI})\beta_S\left(1-\min\left\{1, \frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right\}\right),$$

which simplifies to

$$P_T^G(p) = (1-p)\cdot\begin{cases}(1-\beta_{AI})(1-\beta_S) & \text{if } p \le \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)},\\ (1-\beta_{AI})(1-\beta_S)+(1-\beta_{AI})\beta_S\left(1-\frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right) & \text{otherwise.}\end{cases} \quad (5)$$

For low values of p, the specialist's negative diagnosis acts as a strong adjustment, using AI as a gatekeeper reduces unnecessary treatments. This occurs because negative cases incorrectly flagged as positive by AI are likely to be overturned by the specialist's negative diagnosis when p is low.

The larger the threshold τ_T^G, the broader the set of prior values p for which AI as a gatekeeper reduces unnecessary treatment. Interestingly, the region where gatekeeper AI reduces unnecessary treatment expands with α_S (the specialist's sensitivity) and contracts with β_S (the specialist's specificity). This result contrasts with the intuition that the region should expand as AI's specificity increases. However, initially, the region contracts as β_{AI} increases until β_{AI} reaches a sufficiently high value (approximately $2\beta_S - 1$), after which the region expands as β_{AI} increases further.

4.2. AI as a Second Opinion

Intuitively, using AI as a second opinion after the specialist has made an independent diagnosis offers a potential way to reduce the risk of both missed diagnoses and unnecessary treatments. However, as Propositions 3 and 4 show below, whereas AI as a second opinion can deliver its anticipated benefits in certain scenarios, it may have the opposite effect in others.

Proposition 3 *A threshold τ_M^S exists such that AI as a second opinion increases the probability of missed diagnoses for all $p \in [0, \tau_M^S)$ and reduces (or maintains) the probability of missed diagnoses for all $p \in [\tau_M^S, 1]$:*

$$\tau_M^S = \begin{cases}\frac{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)^2+\alpha_S(1-\alpha_{AI})(1-\beta_{AI})} & \textit{if } \alpha_S \le \alpha_{AI} \textit{ and } 0 \le \frac{1-\beta_{AI}}{1-\beta_S} \le \frac{\alpha_{AI}(1-\alpha_S)^2}{\alpha_S(1-\alpha_{AI})^2}\\ \frac{\alpha_S(1-\beta_S)}{\alpha_S(1-\beta_S)+\alpha_{AI}(1-\alpha_S)} & \textit{if } \alpha_S > \alpha_{AI} \textit{ and } 0 \le \frac{1-\beta_{AI}}{1-\beta_S} \le \frac{\alpha_S}{\alpha_{AI}}\\ \frac{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}}{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}+\sqrt{\alpha_{AI}(1-\alpha_S)^2}} & \textit{otherwise.}\end{cases}$$

REMARK 1. Even when AI performs worse than the specialist in both dimensions ($\alpha_{AI} < \alpha_S$ and $\beta_{AI} < \beta_S$), using AI as a second opinion can reduce the probability of a missed diagnosis. By contrast, AI as a gatekeeper reduces the probability of a missed diagnosis only when $\alpha_{AI} > \alpha_S$.

The probability of a missed diagnosis when the patient only sees the specialist is $P_M^0(p) = p(1-\alpha_S)$. When AI is used as a second opinion, a missed diagnosis can occur in three situations: (i) if both the specialist and AI incorrectly diagnose the patient as negative; (ii) if the specialist correctly diagnoses a positive case as positive but AI misdiagnoses it as negative, and the specialist accepts AI's incorrect diagnosis; or (iii) if the specialist incorrectly diagnoses a positive case as negative and AI correctly identifies it as positive, but the specialist maintains their original negative diagnosis.

The probability of a missed diagnosis when AI is used as a second opinion is

$$\begin{aligned} P_M^S(p) =& p(1-\alpha_S)(1-\alpha_{AI}) + p\alpha_S(1-\alpha_{AI}) \min\left\{1, \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})}\right\} \\ &+ p(1-\alpha_S)\alpha_{AI}\left(1 - \min\left\{1, \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\}\right). \end{aligned} \tag{6}$$

Furthermore,

$$\min\left\{1, \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})}\right\} = \begin{cases} 1 & \text{if } p \le \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} \\ \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})} & \text{if } p > \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} \end{cases}, \text{ and}$$

$$\min\left\{1, \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\} = \begin{cases} \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})} & \text{if } p \le \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})}, \\ 1 & \text{if } p > \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})}. \end{cases}$$

If $\frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} \le \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}$,

$$P_M^S(p) = \begin{cases} [p(1-\alpha_S\alpha_{AI}) - \frac{p^2\alpha_{AI}(1-\alpha_S)^2}{(1-p)(1-\beta_{AI})}] & \text{if } p \le \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})}, \\ p(1-\alpha_{AI}) & \text{if } \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} < p < \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}, \\ [p(1-\alpha_S)(1-\alpha_{AI}) + (1-p)\alpha_S(1-\beta_S)] & \text{if } p \ge \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}. \end{cases} \tag{7}$$

The complementary case (i.e., $\frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} > \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}$) follows similarly:

$$P_M^S(p) = \begin{cases} [p(1-\alpha_S\alpha_{AI}) - \frac{p^2\alpha_{AI}(1-\alpha_S)^2}{(1-p)(1-\beta_{AI})}] & \text{if } p \le \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}, \\ [p(1-\alpha_S) - \frac{p^2\alpha_{AI}(1-\alpha_S)^2}{(1-p)(1-\beta_{AI})} + (1-p)\alpha_S(1-\beta_S)] & \text{if } \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} < p < \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})}, \\ [p(1-\alpha_S)(1-\alpha_{AI}) + (1-p)\alpha_S(1-\beta_S)] & \text{if } p \ge \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})}. \end{cases} \tag{8}$$

For small prior values of p, the strength of the adjustment exceeds that of the anchor, making the adjustment more likely to be accepted. As a result, using AI as a second opinion may increase missed diagnoses when priors are low. However, in cases where the specialist initially misdiagnoses a positive case as negative and AI correctly identifies it as positive, the specialist may still adhere to their original diagnosis. This is particularly likely when the prior is low, as the adjustment from AI has a weaker influence compared to the anchor.

The following proposition outlines the condition under which using AI as a second opinion helps reduce overtreatment:

Proposition 4 *A threshold $\tau_T^S \in (0,1)$ exists such that AI as a second opinion reduces or maintains the probability of unnecessary treatment for all $p \in [0, \tau_T^S]$ and increases the probability of unnecessary treatment for all $p \in (\tau_T^S, 1]$:*

$$\tau_T^S = \begin{cases} \frac{\beta_{AI}(1-\beta_S)}{\beta_{AI}(1-\beta_S)+\beta_S(1-\alpha_S)} & \text{if } \beta_S > \beta_{AI} \text{ and } 0 \le \frac{1-\alpha_{AI}}{1-\alpha_S} \le \frac{\beta_S}{\beta_{AI}} \\ \frac{(1-\beta_S)^2\beta_{AI}}{(1-\beta_S)^2\beta_{AI}+\beta_S(1-\beta_{AI})(1-\alpha_{AI})} & \text{if } \beta_S \le \beta_{AI} \text{ and } 0 \le \frac{1-\alpha_{AI}}{1-\alpha_S} \le \frac{\beta_{AI}(1-\beta_S)^2}{\beta_S(1-\beta_{AI})^2} \\ \frac{\sqrt{(1-\beta_S)^2\beta_{AI}}}{\sqrt{(1-\beta_S)^2\beta_{AI}}+\sqrt{(1-\alpha_{AI})(1-\alpha_S)\beta_S}} & \text{otherwise.} \end{cases}$$

The probability of unnecessary treatment when the patient only sees the specialist is $(1-p)(1-\beta_S)$. When AI is used as a second opinion, unnecessary treatment occurs when a negative case is misdiagnosed as positive and treated unnecessarily, which can happen in the following cases: (i) Both the specialist and AI give a positive diagnosis; (ii) the specialist gives a positive diagnosis and AI gives a negative diagnosis, but the specialist retains their original positive diagnosis; or (iii) the specialist gives a negative diagnosis and AI gives a positive diagnosis, with the specialist accepting AI's positive diagnosis. The overall probability of unnecessary treatment when AI is used as a second opinion is given by

$$\begin{aligned} P_T^S(p) =& (1-p)(1-\beta_S)(1-\beta_{AI}) + (1-p)(1-\beta_S)\beta_{AI}\left(1-\min\left\{1, \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})}\right\}\right) \\ &+ (1-p)\beta_S(1-\beta_{AI})\min\left\{1, \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\}. \end{aligned} \tag{9}$$

If $\frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} \le \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}}$,

$$P_T^S(p) = \begin{cases} (1-p)(1-\beta_S)(1-\beta_{AI}) + p(1-\alpha_S)\beta_S & if \ p \le \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S}, \\ (1-p)(1-\beta_{AI}) & if \ \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} < p < \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}}, \\ (1-p)(1-\beta_{AI}) + (1-p)(1-\beta_S)\beta_{AI} - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} & if \ p \ge \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}}. \end{cases} \tag{10}$$

Alternately, if $\frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} > \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}}$,

$$P_T^S(p) = \begin{cases} (1-p)(1-\beta_S)(1-\beta_{AI}) + p(1-\alpha_S)\beta_S & if \ p \le \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}}, \\ (1-p)(1-\beta_S) + p(1-\alpha_S)\beta_S - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} & if \ \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} < p < \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S}, \\ (1-p)(1-\beta_{AI}) + (1-p)(1-\beta_S)\beta_{AI} - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} & if \ p \ge \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S}. \end{cases} \tag{11}$$

4.3. The Optimal AI Strategy

We consider an altruistic system planner who wants to use AI to minimize missed diagnoses as well as unnecessary treatments. The results in Sections 4.1 and 4.2 show that whether AI is helpful in reducing missed diagnoses or unnecessary treatments depends on the prior probability, p, of whether the patient has the disease. Suppose the cost incurred by a patient due to a missed diagnosis is C_M and the cost incurred due to an unnecessary treatment is C_T, and the system planner wants to minimize the total expected cost of a missed diagnosis and unnecessary treatment for a prior p. For simplicity, assume the prior p captures the prevalence of the disease.[1] We investigate the optimal AI strategy for the system planner by answering the following questions:

1. For what values of p is using AI as a gatekeeper the optimal strategy?
2. For what values of p is using AI as a second opinion the optimal strategy?
3. Is it always optimal to use AI, either as a gatekeeper or as a second opinion, even if it is costless?

For a prior p, the total expected cost of a missed diagnosis and unnecessary treatment under no AI, AI as a gatekeeper, and AI as a second opinion are, respectively,

$$\begin{aligned} C_0(p) &= C_M p(1-\alpha_S) + C_T(1-p)(1-\beta_S), \\ C_G(p) &= C_M P_M^G(p) + C_T P_T^G(p), \text{ and} \\ C_S(p) &= C_M P_M^S(p) + C_T P_T^S(p). \end{aligned} \tag{12}$$

Substituting the values of $P_M^G(p)$ and $P_T^G(p)$ yields

$$C_G(p) = \begin{cases} C_M p(1-\alpha_{AI}\alpha_S) + C_T(1-p)(1-\beta_{AI})(1-\beta_S) & \text{if } p \le p_1 \\ \frac{C_M(1-\alpha_S)(1-\alpha_{AI})p^2 + (C_M\alpha_{AI}+C_T)(1-\beta_{AI})(1-\alpha_S)p(1-p) - C_T\beta_S(1-\beta_{AI})^2(1-p)^2}{p(1-\alpha_S)} & \text{if } p > p_1 \end{cases}, \tag{13}$$

where

$$p_1 = \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)}.$$

We make the following assumption:

Assumption 1 $(1-\alpha_{AI})(1-\beta_{AI}) \le (1-\alpha_S)(1-\beta_S)$

Under Assumption 1, $p_1 \le p_2$, and the total expected cost under AI as a second opinion is

$$C_S(p) = \begin{cases} \frac{(1-p)^2[C_T(1-\beta_S)(1-\beta_{AI})^2] + p(1-p)[C_M(1-\beta_{AI})(1-\alpha_{AI}\alpha_S) + C_T\beta_S(1-\alpha_S)(1-\beta_{AI})] - p^2\alpha_{AI}(1-\alpha_S)^2 C_M}{(1-p)(1-\beta_{AI})} & \text{if } p \le p_1 \\ C_M p(1-\alpha_{AI}) + C_T(1-p)(1-\beta_{AI}) & \text{if } p_1 < p < p_2 \\ \frac{C_M(1-\alpha_S)(1-\alpha_{AI})^2 p^2 + (1-\alpha_{AI})(C_M\alpha_S(1-\beta_S) + C_T(1-\beta_S\beta_{AI}))p(1-p) - C_T\beta_{AI}(1-\beta_S)^2(1-p)^2}{p(1-\alpha_{AI})} & \text{if } p \ge p_2 \end{cases}, \tag{14}$$

[1] To generate sharp managerial insights, we assume both the specialist and AI to be costless. Equivalently, we assume the system designed is solely concerned about the well-being of the patient rather than the costs of consultation.

where

$$p_1 = \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)} \text{ and } p_2 = \frac{(1-\beta_S)}{(1-\beta_S)+(1-\alpha_{AI})}.$$

Gatekeeper AI vs. No AI. First, we investigate the optimal strategy with using AI as a gatekeeper. Specifically, we find priors where using AI as a gatekeeper has a lower total expected cost of a missed diagnosis and unnecessary treatment than when AI is not used. Substituting the expressions from eqs. (12) and (13) yields

$$C_G(p) - C_0(p) = \begin{cases} g_1(p) & \text{if } p \le p_1, \\ g_2(p) & \text{if } p > p_1. \end{cases}$$

where

$$\begin{aligned} g_1(p) &= C_M p(1-\alpha_{AI})\alpha_S - C_T(1-p)\beta_{AI}(1-\beta_S), \text{ and} \\ g_2(p) &= C_M p(\alpha_S - \alpha_{AI}) + C_M(1-p)\alpha_{AI}(1-\beta_{AI}) \\ &\quad - C_T(1-p)\beta_{AI}(1-\beta_S) + C_T(1-p)(1-\beta_{AI})\beta_S\left(1 - \frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right). \end{aligned}$$

When the prior is small ($p \le p_1$), the gatekeeper AI's positive signal has no anchoring effect. The net expected cost difference between using AI as a gatekeeper and not using AI ($g_1(p)$) is driven by two factors: the expected increase in the cost of a missed diagnosis with AI and the expected reduction in unnecessary treatments with AI. The increase in missed diagnosis cost is captured by $(1-\alpha_{AI})\alpha_S$, representing the probability that AI misses a positive case that the specialist would have caught. The reduction in unnecessary treatment cost is captured by $\beta_{AI}(1-\beta_S)$, representing the probability that AI correctly screens out a negative case that the specialist would have falsely declared as positive. Further,

$$g_1(p) \le 0 \iff p \le \tau_G^a = \frac{C_T\beta_{AI}(1-\beta_S)}{C_M(1-\alpha_{AI})\alpha_S + C_T\beta_{AI}(1-\beta_S)}.$$

When the prior is sufficiently high ($p > p_1$), the anchoring effect also comes into play. Anchoring to the positive signal from the gatekeeper AI can, on one hand, help reduce missed diagnoses by influencing the specialist's decision. On the other hand, it can also lead to an increase in unnecessary treatments.

Let $D = (C_M\alpha_{AI}(1-\beta_{AI}) + C_T(\beta_S - \beta_{AI}))^2(1-\alpha_S)^2 + 4C_M C_T\beta_S(1-\beta_{AI})^2(\alpha_S - \alpha_{AI})(1-\alpha_S)$. Suppose $\alpha_S \ne \alpha_{AI}$. For $D \ge 0$,

$$g_2(p) \le 0 \iff (\alpha_S - \alpha_{AI})(p - \tau_G^b)(p - \tau_G^c) \le 0,$$

where

$$\tau_G^b = \frac{\sqrt{D} - (1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI}) + C_T(\beta_S - \beta_{AI}))}{\sqrt{D} - (1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI}) + C_T(\beta_S - \beta_{AI})) + 2C_M(\alpha_S - \alpha_{AI})(1-\alpha_S)}$$

$$\tau_G^c = \frac{\sqrt{D} + (C_M\alpha_{AI}(1-\beta_{AI}) + C_T(\beta_S - \beta_{AI}))(1-\alpha_S)}{\sqrt{D} + (C_M\alpha_{AI}(1-\beta_{AI}) + C_T(\beta_S - \beta_{AI}) + 2C_M(\alpha_{AI} - \alpha_S))(1-\alpha_S)}. \tag{15}$$

For $D \geq 0$ and $\alpha_S < \alpha_{AI}$, the use of AI as a gatekeeper reduces the total cost for low and high prior probabilities. Conversely, for $D \geq 0$ and $\alpha_S > \alpha_{AI}$, AI as a gatekeeper reduces the total cost for intermediate priors. When $D < 0$ and $\alpha_S < \alpha_{AI}$, $g_2(p) < 0$ for all $p \in [0,1]$, meaning AI as a gatekeeper reduces the total cost across all prior probabilities. By contrast, for $D < 0$ and $\alpha_S > \alpha_{AI}$, $g_2(p) > 0$ for all $p \in [0,1]$, indicating AI as a gatekeeper increases the total cost for all priors.

Proposition 5 (Gatekeeper AI vs. No AI) *Suppose $\beta_S \geq \beta_{AI}$. The following conditions determine when AI as a gatekeeper yields a lower or equal total expected cost of a missed diagnosis and unnecessary treatment compared to the case in which AI is not used:*

1. *If $\alpha_S < \alpha_{AI}$ and $\frac{C_M}{C_T} \geq \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$, $D \geq 0$, and $0 \leq \tau_G^a \leq \tau_G^c \leq 1$, and the total expected cost under AI as a gatekeeper is less than or equal to that under no AI for all priors $p \in [0, \tau_G^a] \cup [\tau_G^c, 1]$.*
2. *If $\alpha_S < \alpha_{AI}$ and $D \geq 0$, the total expected cost under AI as a gatekeeper is less than or equal to that under no AI for all priors $p \in [0, \tau_G^b] \cup [\tau_G^c, 1]$, provided that either $\alpha_{AI} + \alpha_{AI}\alpha_S - 2\alpha_S \leq 0$ or $\alpha_{AI} + \alpha_{AI}\alpha_S - 2\alpha_S > 0$ and $\frac{C_M}{C_T} < \min\left\{\frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}, \frac{(\beta_S-\beta_{AI})(1-\alpha_S)}{\alpha_{AI}+\alpha_{AI}\alpha_S-2\alpha_S}\right\}$. Under these conditions, the thresholds $0 \leq \tau_G^b \leq \tau_G^c \leq 1$ ensure that AI as a gatekeeper is cost-effective within the specified prior intervals.*
3. *If $\alpha_S < \alpha_{AI}$, $D \geq 0$, $(\alpha_{AI} + \alpha_{AI}\alpha_S - 2\alpha_S) > 0$, and $\frac{(\beta_S-\beta_{AI})(1-\alpha_S)}{\alpha_{AI}+\alpha_{AI}\alpha_S-2\alpha_S} \leq \frac{C_M}{C_T} \leq \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$, or if $\alpha_S < \alpha_{AI}$, $D < 0$, and $\frac{C_M}{C_T} < \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$, then the total expected cost under AI as a gatekeeper is less than or equal to that under no AI for all priors $p \in [0,1]$.*
4. *If $\alpha_S > \alpha_{AI} > 0$, then $D > 0$ and $\tau_G^c \leq 0 \leq \tau_G^b \leq 1$, and $\frac{C_M}{C_T} < \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$, then the total expected cost under AI as a gatekeeper is less than or equal to that under no AI for all priors $p \in [0, \tau_G^b]$. Otherwise, if $\frac{C_M}{C_T} \geq \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$, then $0 \leq \tau_G^a \leq 1$, and the total expected cost under AI as a gatekeeper is less than or equal to that under no AI for all priors $p \in [0, \tau_G^a]$.*

Propositions 1 and 2 show AI as a gatekeeper is useful for high priors to reduce missed diagnoses (only if $\alpha_{AI} > \alpha_S$) and for low priors to reduce unnecessary treatment. Therefore, optimizing for both missed diagnosis and unnecessary treatment leads to using as a gatekeeper for low and high priors (for $\alpha_{AI} > \alpha_S$), as shown in Proposition 5 (see Table A4 for the full characterization of the results of Proposition 5).

Intriguingly, even when AI is both less sensitive and less specific than the specialist, using AI as a gatekeeper to reduce unnecessary treatments for low priors may still be optimal. On the other hand, if AI's sensitivity is significantly higher than that of the specialist, using AI as a gatekeeper becomes optimal for all priors.

REMARK 2. In the absence of anchoring, where only the adjustment (the second diagnosis) is retained, the prior threshold below which AI as a gatekeeper has a lower total expected cost of a missed diagnosis and unnecessary treatment than no AI is given by $\tau_G^a = \frac{C_T\beta_{AI}(1-\beta_S)}{C_M(1-\alpha_{AI})\alpha_S + C_T\beta_{AI}(1-\beta_S)}$. The value of τ_G^a increases with $C_T\beta_{AI}(1-\beta_S)$, which represents the savings in the cost of unnecessary treatment when AI correctly provides a negative diagnosis (with probability β_{AI}) but the specialist gives a false-positive diagnosis (with probability $1-\beta_S$). As this savings grows, the range of priors for which AI as a gatekeeper outperforms no AI widens. Conversely, τ_G^a decreases with $C_M(1-\alpha_{AI})\alpha_S$, which captures the increased cost of a missed diagnosis when AI incorrectly provides a negative diagnosis (with probability $1-\alpha_{AI}$) while the specialist correctly identifies a positive diagnosis (with probability α_S). A higher value of this term narrows the range of priors for which AI as a gatekeeper dominates no AI.

The prior threshold above which the anchoring effect exists when AI is used as a gatekeeper is given by $p_1 = \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S}$. This threshold decreases as β_{AI} increases, indicating that as AI becomes more specific, the range of priors where the specialist is anchored to AI's positive diagnosis expands. Conversely, p_1 increases with α_S, showing that as the specialist becomes more sensitive, the range of priors where the specialist is anchored to AI's positive diagnosis contracts.

The condition $\frac{C_M}{C_T} \geq \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$ is equivalent to the requirement that $\tau_G^a \leq p_1$. This finding implies that when the cost of a missed diagnosis is sufficiently higher than the cost of unnecessary treatment, a range of priors exists, including $[\tau_G^a, p_1]$, where the anchoring effect of AI's positive diagnosis does not manifest, and no AI outperforms AI as a gatekeeper. Similarly, for $D > 0$ and $\alpha_S < \alpha_{AI}$, the condition $C_M(\alpha_{AI} + \alpha_{AI}\alpha_S - 2\alpha_S) < C_T(\beta_S - \beta_{AI})(1-\alpha_S)$ guarantees the existence of a range of priors where, even when the anchoring effect is present, no AI dominates AI as a gatekeeper. This condition holds when the difference in specificities $(\beta_S - \beta_{AI})$ is sufficiently large or when the difference in sensitivities $(\alpha_{AI} - \alpha_S)$ is sufficiently small.

Lemma 1 *The expected cost of a missed diagnosis under AI as a gatekeeper is greater than or equal to the expected cost of a missed diagnosis under AI as a second opinion, that is, for* $C_T = 0$, $C_G(p) \geq C_S(p)$ *for all priors* $p \in [0,1]$.

Using AI as a second opinion provides more opportunity to reduce missed diagnoses by allowing the specialist to overrule a negative diagnosis by AI.

Lemma 2 *The expected cost of unnecessary treatment under AI as a second opinion is greater than or equal to the expected cost of unnecessary treatment under AI as a gatekeeper for all priors* $p \in [0,1]$.

Using AI as a gatekeeper offers greater potential to reduce unnecessary treatments by effectively screening out negative cases upfront.

Given the importance of reducing missed diagnoses for high priors and minimizing unnecessary treatments for low priors, the above lemmas suggest it may be optimal to employ AI as a gatekeeper for low priors and as a second opinion for high priors. The next result formalizes this intuition.

Proposition 6 *If* $\beta_S > \beta_{AI}$, $(1-\alpha_{AI})(1-\beta_{AI}) \le (1-\alpha_S)(1-\beta_S)$, $\beta_{AI} < \alpha_S + \alpha_{AI}$ *and* $C_M \ge C_T$, *a threshold* τ_S^G *exists such that the total expected cost of a missed diagnosis and unnecessary treatment under AI as a second opinion is less than or equal to that under AI as a gatekeeper for all priors* p *in* $[\tau_S^G, 1]$:

$$\tau_S^G = \begin{cases} \frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI})+C_M\alpha_{AI}(1-\alpha_S)} & \text{for } \frac{C_M}{C_T} > \frac{\beta_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)(1-\beta_S)}, \\ \frac{\left[C_T\beta_{AI}(1-\beta_S)+C_M(\alpha_S(1-\beta_S)-\alpha_{AI}(1-\beta_{AI})\right](1-\alpha_S)+\sqrt{D'}}{\left[C_T\beta_{AI}(1-\beta_S)+C_M(\alpha_S(1-\beta_S)-\alpha_{AI}(1-\beta_{AI})+2\alpha_S(1-\alpha_{AI}))\right](1-\alpha_S)+\sqrt{D'}} & \text{for } \frac{C_M}{C_T} \le \frac{\beta_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)(1-\beta_S)} \end{cases}$$

where $D' = \left[(C_T\beta_{AI} + C_M\alpha_S)(1-\beta_S) - C_M\alpha_{AI}(1-\beta_{AI})\right]^2(1-\alpha_S)^2 + 4C_MC_T\alpha_S(1-\alpha_S)\left[\beta_S(1-\beta_{AI})^2(1-\alpha_{AI}) - \beta_{AI}(1-\beta_S)^2(1-\alpha_S)\right]$ *such that* $\frac{C_M}{C_T} \le \frac{\beta_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)(1-\beta_S)} \implies D' \ge 0$. *In both cases* $0 \le \tau_S^G \le 1$.

Table A5 provides a full characterization of the results of Proposition 6. In the proposition, the assumption $(1-\alpha_{AI})(1-\beta_{AI}) \le (1-\alpha_S)(1-\beta_S)$ ensures AI is not worse than the specialist in both sensitivity and specificity. For instance, if $\beta_S \ge \beta_{AI}$, it must follow that $\alpha_S \le \alpha_{AI}$. This condition also implies the prior threshold $p_1 = \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S}$, above which the anchoring effect of AI's positive diagnosis exists, is less than or equal to the prior threshold $p_2 = \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}}$, which represents the threshold above which the anchoring effect of the specialist's positive diagnosis exists. Consequently, when making the final decision, the specialist is anchored to AI's initial positive diagnosis over a wider range of priors than they are to their own initial positive diagnosis.

In the absence of the anchoring effect of the specialist's positive diagnosis, AI as a second opinion outperforms AI as a gatekeeper for all priors greater than the threshold $\tau_S^G = \frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI})+C_M\alpha_{AI}(1-\alpha_S)}$. The value of τ_S^G increases with $C_T\beta_S(1-\beta_{AI})$, which represents the increased cost of unnecessary treatment caused by AI when the specialist correctly provides a negative diagnosis (with probability β_S) but AI incorrectly provides a positive diagnosis (with probability $1-\beta_{AI}$). As this cost grows, the range of priors where AI as a second opinion outperforms AI as a gatekeeper narrows. Conversely, τ_S^G decreases with $C_M\alpha_{AI}(1-\alpha_S)$, which reflects the reduction in the cost of a missed diagnosis due to AI when the specialist wrongly provides a negative diagnosis (with probability $1-\alpha_S$) but AI correctly provides a positive diagnosis (with probability α_{AI}). A higher value of this term widens the range of priors where AI as a second opinion outperforms AI as a gatekeeper.

The condition $\frac{C_M}{C_T} > \frac{\beta_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)(1-\beta_S)}$ is equivalent to $\tau_S^G < p_2$. This indicates the existence of a range of priors, specifically $[\tau_S^G, p_2]$, where AI as a second opinion outperforms AI as a gatekeeper, even in the absence of the anchoring effect of the specialist's positive diagnosis. Conversely, if $\frac{C_M}{C_T} \le \frac{\beta_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)(1-\beta_S)}$, a range of priors exists where AI as a gatekeeper outperforms AI as a second opinion, even when the anchoring effect of the specialist's positive diagnosis is present.

The next result, derived from Propositions 5 and 6, establishes that when the cost of a missed diagnosis (C_M) is significantly higher than the cost of unnecessary treatment (C_T), intermediate priors (i.e., high-uncertainty cases) exist where the no-AI option (i.e., not using AI) results in a lower expected cost of missed diagnoses and unnecessary treatments compared to using AI as a gatekeeper or as a second opinion. By contrast, when C_M and C_T are comparable, it is always optimal to use AI—as a gatekeeper for low priors and as a second opinion for high priors.

Proposition 7 *Consider the values D, τ_G^a, τ_G^b, τ_G^c, and τ_S^G as defined in Propositions 5 and 6. Suppose the conditions $\beta_S > \beta_{AI}$, $\beta_{AI} < \alpha_S + \alpha_{AI}$, and $(1-\alpha_{AI})(1-\beta_{AI}) \le (1-\alpha_S)(1-\beta_S)$ hold. The following strategies describe the optimal use of AI in clinical decision-making:*

1. *Gatekeeper AI for low priors, no AI for intermediate priors, and second-opinion AI for high priors: If $\alpha_{AI}\beta_{AI}(1-\alpha_S)(1-\beta_S) < \alpha_S\beta_S(1-\alpha_{AI})(1-\beta_{AI})$ and*

$$\frac{C_M}{C_T} > \max\left\{1, \frac{\beta_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)(1-\beta_S)}, \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}, \frac{\beta_S}{\alpha_{AI}}\right\},$$

   where $\tau_G^a < \tau_S^G \le \tau_G^c$, the optimal strategy is:
   - *For all priors $p \in [0, \tau_G^a]$, AI as a gatekeeper is optimal.*
   - *For all priors $p \in [\tau_G^a, \tau_S^G]$, no AI is optimal.*
   - *For all priors $p \in [\tau_S^G, 1]$, AI as a second opinion is optimal.*

2. *Using AI is optimal for all priors—as a gatekeeper for low priors and as a second opinion for high priors: If $(\alpha_{AI} + \alpha_{AI}\alpha_S - 2\alpha_S) > 0$, $D \ge 0$, and*

$$\max\left\{1, \frac{(\beta_S - \beta_{AI})(1-\alpha_S)}{(\alpha_{AI} + \alpha_{AI}\alpha_S - 2\alpha_S)}\right\} \le \frac{C_M}{C_T} \le \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})},$$

   then for all priors $p \in [0, \tau_S^G]$, AI as a gatekeeper is optimal, and for all priors $p \in [\tau_S^G, 1]$, AI as a second opinion is optimal.

In Proposition 7, the assumption $\alpha_{AI}\beta_{AI}(1-\alpha_S)(1-\beta_S) < \alpha_S\beta_S(1-\alpha_{AI})(1-\beta_{AI})$ implies that, in the absence of anchoring effects, the threshold prior above which AI as a second opinion dominates AI as a gatekeeper, given by $\frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI})+C_M\alpha_{AI}(1-\alpha_S)}$, is higher than the threshold prior above which no AI dominates AI as a gatekeeper, given by $\frac{C_T\beta_{AI}(1-\beta_S)}{C_M(1-\alpha_{AI})\alpha_S+C_T\beta_{AI}(1-\beta_S)}$. This relationship indicates that, in the absence of anchoring effects, the no-AI option dominates both AI as a gatekeeper and AI as a second opinion for prior values lying between these two thresholds.

The restriction on $\frac{C_M}{C_T}$ in Proposition 7(1) ensures the two thresholds—the one above which AI as a second opinion dominates AI as a gatekeeper and the one above which no AI dominates AI as a gatekeeper—are smaller than the prior thresholds where the anchoring effects of positive diagnoses by the specialist and AI emerge, respectively. This condition guarantees the existence of a range of priors where no AI is the optimal strategy. By contrast, the restriction on $\frac{C_M}{C_T}$ in Proposition 7(2) implies that AI as a gatekeeper dominates no AI for all priors, and AI as a second opinion dominates AI as a gatekeeper for all priors above a certain threshold.

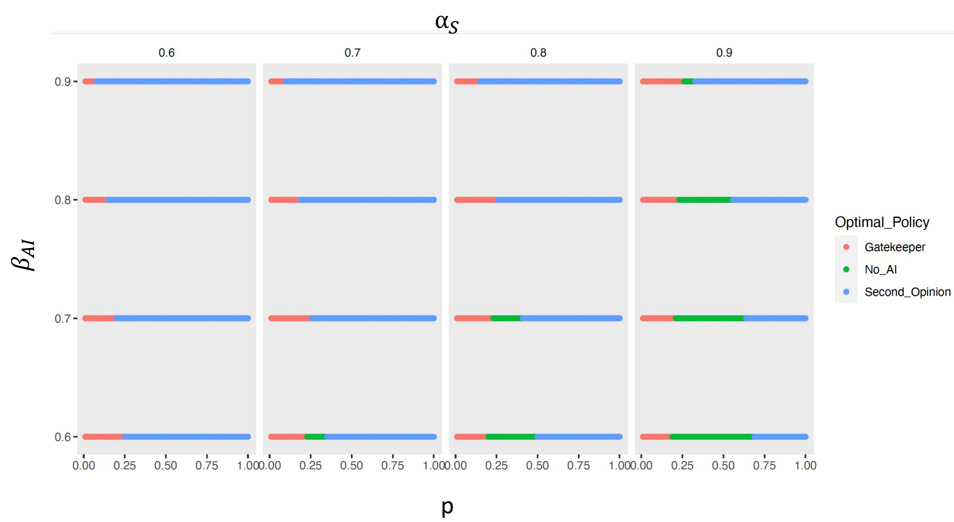

Figure 1 **Optimal policy for different values of p and β_{AI} for $C_T = 50$ and $C_M = 150$. The numbers on the top of the graphs are the different values of α_S.**

Figure 1 illustrates how the optimal strategy for using AI varies with the prior p of a patient having the disease, assuming $\alpha_{AI} = \beta_S = 0.95$, while adjusting α_S (top labels of the graphs) and β_{AI} (y-axis). When the specialist's sensitivity is low ($\alpha_S = 0.6$), AI as a gatekeeper is optimal for lower priors p and transitions to AI as a second opinion as the prior increases, even when AI's specificity is relatively low ($\beta_{AI} = 0.6$). As β_{AI} increases, the range of priors favoring AI as a second opinion expands, since the primary disadvantage of using AI as a second opinion—its higher risk of unnecessary treatments relative to AI as a gatekeeper (see Lemma 2)—diminishes with a higher β_{AI}. For medium or high α_S, the optimal strategy involves using AI as a gatekeeper for low priors, avoiding AI for intermediate priors, and adopting AI as a second opinion for higher priors. The range of priors favoring no AI expands when β_{AI} is low, as the lower specificity increases unnecessary treatments. As the specialist's sensitivity (α_S) increases, AI as a gatekeeper becomes optimal over a broader range of priors, because the primary drawback of using AI as a gatekeeper—its higher expected cost of missed diagnoses relative to AI as a second opinion (see Lemma 1)—is mitigated by a higher α_S. Furthermore, the range of priors where no AI is optimal expands with lower AI specificity (β_{AI}) or higher specialist sensitivity (α_S).

An example satisfying the conditions of Proposition 7(1) is when $\alpha_{AI} = 0.97$, $\beta_{AI} = 0.62$, $\alpha_S = 0.77$, $\beta_S = 0.95$, $C_M = 150$, and $C_T = 50$. In this case, the thresholds are $\tau_G^a = 0.31$, $\tau_S^G = 0.35$, and $\tau_G^c = 0.65$. For priors $p \in [0.31, 0.35]$, the no-AI strategy is optimal.

5. Pathfinder: Data-Driven Clinical Workflow Optimization

In this section, we utilize the "Pathfinder" framework to optimize patient pathways based on individualized risk assessments. We demonstrate the practical implementation of this system using the PAPILA dataset (Kovalyk et al. 2022), which includes data from 210 glaucoma patients. This dataset contains structured clinical information—such as age, gender, and various eye-specific measurements (e.g., refractive error, astigmatism, and lens status)—as well as high-resolution unstructured data in the form of retinal fundus images for both eyes (Figure 2). The PAPILA dataset is suitable for developing convolutional neural network (CNN)-based AI models that predict a patient's risk profile, enabling personalized clinical decision-making.

(a) Right eye of Patient #68

(b) Left eye of Patient #68

Figure 2 **Fundus images from the PAPILA dataset for both eyes of Patient #68. Source: PAPILA dataset (https://figshare.com/articles/dataset/PAPILA/14798004?file=35013982)**

5.1. Data and Prior Prediction

Our dataset includes both clinical measurements and diagnostic outcomes for a total of 210 patients, of whom 47 are confirmed glaucoma cases (i.e., Diagnosis = 1 in at least one eye) and 163 are confirmed negative cases (i.e., Diagnosis = 0 in both eyes).

We train a probit model to predict the prior probability that a patient has glaucoma based on readily available clinical features: age, gender, diopter values of left and right eyes, astigmatism, and lens status (phakic or pseudophakic) for both eyes. Figure 3 presents the summary of the probit model results, which show significant predictive power with an area under the ROC curve (AUC) of 0.81.

Probit Regression Results

Dep. Variable:	Diagnosis	No. Observations:	210
Model:	Probit	Df Residuals:	199
Method:	MLE	Df Model:	10
Date:	Sun, 24 Mar 2024	Pseudo R-squ.:	0.2236
Time:	15:46:20	Log-Likelihood:	-86.692
converged:	True	LL-Null:	-111.65
Covariance Type:	nonrobust	LLR p-value:	2.756e-07

	coef	std err	z	P>|z|	[0.025	0.975]
const	-4.9165	0.890	-5.524	0.000	-6.661	-3.172
Age	0.0670	0.012	5.456	0.000	0.043	0.091
Gender	-0.4583	0.236	-1.938	0.053	-0.922	0.005
dioptre_1_OD	-0.0085	0.093	-0.091	0.928	-0.191	0.174
dioptre_2_OD	0.1611	0.147	1.098	0.272	-0.127	0.449
astigmatism_OD	-0.0008	0.003	-0.273	0.784	-0.007	0.005
Phakic/Pseudophakic_OD	-0.2965	0.664	-0.446	0.655	-1.599	1.006
dioptre_1_OC	-0.0866	0.080	-1.079	0.281	-0.244	0.071
dioptre_2_OC	-0.0035	0.026	-0.135	0.892	-0.054	0.047
astigmatism_OC	0.0023	0.003	0.690	0.490	-0.004	0.009
Phakic/Pseudophakic_OC	0.6520	0.676	0.964	0.335	-0.674	1.978

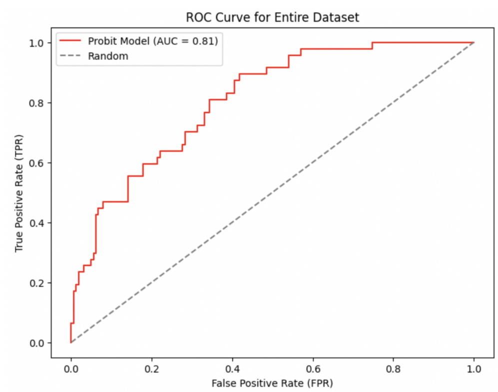

Figure 3 **Summary of results of the probit regression model to predict the prior probability of glaucoma**

5.2. Simulating the Pathfinder in a Clinical Workflow

We demonstrate the implementation of the Pathfinder system by assigning each patient to their optimal path based on the prior probability predicted by the probit model. Following Lim et al. (2023), we assume specialists have high specificity (95%) but low sensitivity (60%), whereas AI systems typically have high sensitivity (85–95%) but lower specificity (~~85~~70–90%).

In our simulation, each patient is assigned to an optimal clinical pathway—AI as a gatekeeper, AI as a second opinion, or no AI—based on expected costs, which account for the costs of a missed diagnosis (C_M) and unnecessary treatment (C_T). The simulation assesses cost effectiveness by predicting the prior probability of glaucoma and comparing the expected total costs across the three pathways.

5.3. Results: Cost Reduction from Using AI

Figure 4 summarizes the cost analysis for different configurations of AI sensitivity and specificity, alongside the cost savings achieved through personalized pathways using the Pathfinder system.

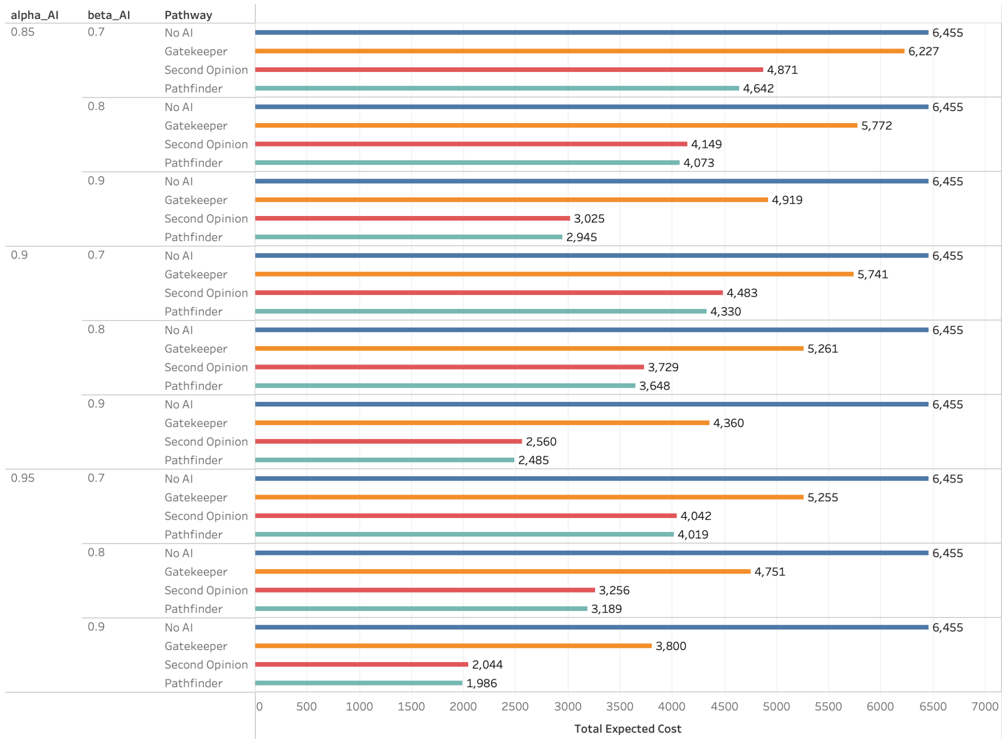

Figure 4 **Total expected cost of a missed diagnosis and unnecessary treatment for $C_M = 150$, $C_T = 50$, $\alpha_S = 0.6$, and $\beta_S = 0.95$ under different design configurations, with percentage cost reduction from personalization over the best standard pathway.**

When C_M is three times higher than C_T, using AI as a second opinion results in substantial cost reductions, particularly when AI sensitivity and specificity are both high. For lower cost ratios, using AI as a gatekeeper proves more effective in minimizing total costs.

5.4. Managerial Insights

The results of our analysis offer key insights for integrating AI into clinical decision-making, particularly in the context of glaucoma screening. These insights hinge on the interaction between specialist expertise, AI performance, and the relative costs of misdiagnosis versus unnecessary treatment.

First, when specialists have relatively low sensitivity, our findings suggest using AI—whether as a gatekeeper or a second opinion—leads to lower overall costs than scenarios without AI. This result holds even if AI has average quality, characterized by medium sensitivity and low specificity. The choice between AI as a gatekeeper or as a second opinion depends on the trade-off between the costs of missed diagnoses and unnecessary treatments. When the cost of a missed diagnosis, C_M, is significantly higher than the cost of unnecessary treatment, C_T, AI as a second opinion consistently results in lower expected costs, regardless of AI's sensitivity or specificity. In such

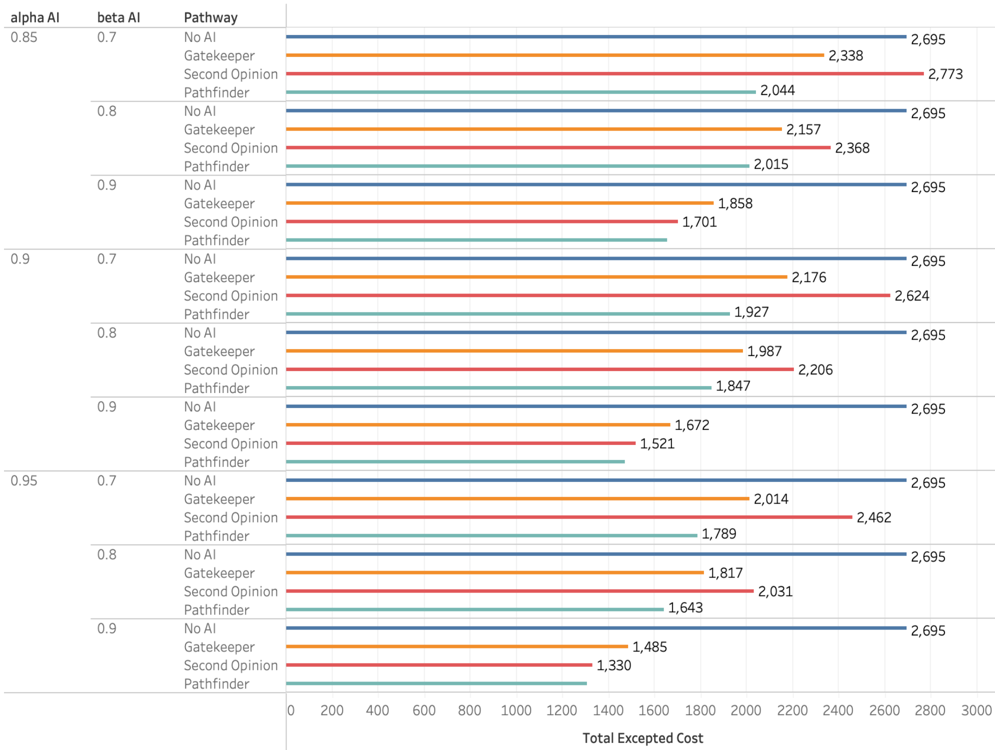

Figure 5 **Total expected cost of a missed diagnosis and unnecessary treatment for $C_M = 50$, $C_T = 50$, $\alpha_S = 0.6$, and $\beta_S = 0.95$ under different design configurations, with percentage cost reduction from personalization over the best standard pathway.**

cases, AI's primary role is to reduce false negatives, making it beneficial to use AI as an additional layer of validation after the specialist.

On the other hand, when C_M and C_T are similar, using AI as a gatekeeper is generally the optimal strategy, especially when AI has medium or high sensitivity. Under these conditions, AI can effectively filter out low-risk cases, reducing the specialist's workload and minimizing unnecessary interventions. Interestingly, AI as a gatekeeper achieves greater cost reductions than AI as a second opinion when AI specificity is low, because it efficiently screens a large patient pool.

Our analysis shows the Pathfinder system, which personalizes the clinical pathway for each patient based on predicted risk, consistently reduces costs compared to the best uniform approach across all scenarios. These cost savings are most pronounced when AI is of average quality—neither highly sensitive nor highly specific—because personalized pathways tailor the combination of AI and specialist interventions to optimize outcomes for each patient.

For "highly skilled" specialists with high sensitivity (resulting in fewer missed diagnoses), uniformly applying AI to all patients may be less effective, particularly when AI sensitivity is moderate or specificity is low. When C_M is much higher than C_T, AI as a second opinion remains advantageous, provided AI meets a sufficient quality threshold. In such cases, AI as a second opinion outperforms AI as a gatekeeper when AI has medium sensitivity and high specificity, striking a balance between identifying true positives and avoiding false positives. Conversely, when C_M and C_T are comparable, AI as a gatekeeper proves more effective, especially when AI sensitivity is high, as it reduces unnecessary specialist referrals while maintaining its ability to identify high-risk cases.

Moreover, the Pathfinder system's personalized approach consistently reduces costs compared to any uniform strategy, even when using AI as a gatekeeper or a second opinion for all patients is not optimal. The value of personalization is especially evident when AI quality is moderate, requiring a nuanced approach that considers both the specialist's skill and AI limitations. Our results indicate that personalization yields the highest cost reductions when AI has average quality, making the Pathfinder system's personalization component critical in balancing the trade-offs between misdiagnosis and unnecessary treatments.

In summary, these insights underscore the importance of purposefully integrating AI into clinical workflows, tailoring AI's role based on both AI system performance and specialist expertise, as well as the relative costs of diagnostic outcomes. Data-driven patient pathways enable more efficient allocation of clinical resources, improving patient outcomes while optimizing overall costs.

6. Conclusion

This paper explores how AI can be integrated into healthcare systems to improve patient outcomes, in line with the U.N.'s Sustainable Development Goal 3 on "good health and well-being." We show AI can help reduce missed diagnoses and unnecessary care, particularly for patients at the high and low ends of the risk spectrum. Yet, for intermediate-risk patients—for whom diagnostic uncertainty is greatest—the use of AI may lead to worse patient outcomes. Our finding points to the need to tailor AI integration to individual risk profiles, rather than relying on a one-size-fits-all approach.

Our paper models a physician as a decision-maker influenced by anchoring bias, which affects patient-pathway decisions when combined with AI. Our analysis shows the sequencing of AI and human expertise plays a role in influencing patient outcomes. Specifically, using AI as a gatekeeper can reduce missed diagnoses but may increase unnecessary treatments, due to specialists' anchoring on initial AI results. On the other hand, using AI as a second opinion can reduce unnecessary treatments but potentially lead to more missed diagnoses. These findings are in line with strategic intertemporal choice in supply chain management (e.g., Lee and Tang 1997).

Using glaucoma diagnosis as a case study, we quantify the cost implications of integrating AI pathways tailored to patient characteristics. When specialists possess lower diagnostic skill, even uniform AI pathways—whether as a gatekeeper or a second opinion—can reduce the total expected cost of care. However, pathways tailored to individual patient characteristics yield greater efficiency, particularly when the diagnostic quality of AI is moderate. By contrast, for highly skilled specialists, uniform AI pathways may increase costs, especially when AI exhibits lower specificity. In these scenarios, personalized pathways demonstrate consistent value, ensuring AI augments—rather than undermines—clinical outcomes.

In conclusion, personalized patient pathways for AI-augmented healthcare offer a promising avenue to improving healthcare delivery. Drawing parallels with principles of supply chain management, AI-driven personalized pathways can optimize resource allocation and care delivery by tailoring interventions to patient-specific needs (Dai). By aligning these strategies with the U.N.'s SDGs, AI can emerge as a transformative tool, driving improvements in patient outcomes across diverse healthcare settings.

References

Abràmoff MD, Dai T, Zou J (2024) Scaling adoption of medical AI — reimbursement from value-based care and fee-for-service perspectives. *NEJM AI* 1(5):AIpc2400083.

Abràmoff MD, Whitestone N, Patnaik JL, Rich E, Ahmed M, Husain L, Hassan MY, Tanjil MSH, Weitzman D, Dai T, Wagner BD, Cherwek DH, Congdon N, Islam K (2023) Autonomous artificial intelligence increases real-world specialist clinic productivity in a cluster-randomized trial. *npj Digital Medicine* 6(1).

Adida E, Dai T (2024) Impact of physician payment scheme on diagnostic effort and testing. *Management Science* 70(8):5408–5425.

Agarwal N, Moehring A, Rajpurkar P, Salz T (2024) Combining human expertise with artificial intelligence: Experimental evidence from radiology. *National Bureau of Economic Research Working Paper* .

Ahmed M, Dai T, Channa R, Abràmoff MD, Lehmann HP, Wolf RM (2024) Cost-effectiveness of AI for pediatric diabetic eye exams from a health system perspective, accepted for publication in *npj Digital Medicine*.

American College of Radiology (2021) Integrating AI into the Clinical Workflow; Case.

Bach AKP, Nørgaard TM, Brok JC, van Berkel N (2023) "If I had all the time in the world": Ophthalmologists' perceptions of anchoring bias mitigation in clinical AI support (New York, NY, USA: Association for Computing Machinery), URL http://dx.doi.org/10.1145/3544548.3581513.

Balakrishnan M, Ferreira K, Tong J (2022) Improving human-algorithm collaboration: Causes and mitigation of over- and under-adherence, working Paper.

CDC (2024) National diabetes statistics report. Centers for Disease Control and Prevention website, URL https://www.cdc.gov/diabetes/php/data-research/index.html, accessed: December 13, 2024.

Chen L, Kim SH, Lee HL (2021) Vehicle maintenance contracting in developing economies: The role of social enterprise. *Manufacturing & Service Operations Management* 23(6):1651–1668.

Cui R, Li M, Zhang S (2022) AI and procurement. *Manufacturing & Service Operations Management* 24(2):691–706.

Dai T, Abràmoff MD (2023) Incorporating artificial intelligence into healthcare workflows: Models and insights. *Tutorials in Operations Research: Advancing the Frontiers of OR/MS: From Methodologies to Applications*, 133–155 (INFORMS).

Dai T, Singh S (2020) Conspicuous by its absence: Diagnostic expert testing under uncertainty. *Marketing Science* 39(3):540–563.

Dai T, Singh S (2024) Artificial intelligence on call: The physician's decision of whether to use AI in clinical practice. Working paper, Johns Hopkins University.

Dai T, Tayur S (2020) OM Forum—Healthcare operations management: A snapshot of emerging research. *Manufacturing & Service Operations Management* 22(5):869–887.

Dai T, Tayur S (2022) Designing AI-augmented healthcare delivery systems for physician buy-in and patient acceptance. *Production and Operations Management* 31(12):4443–4451.

Das Gupta A, Karmarkar US, Roels G (2016) The design of experiential services with acclimation and memory decay: Optimal sequence and duration. *Management Science* 62(5):1278–1296.

de Véricourt F, Gurkan H (2023) Is your machine better than you? You may never know. *Management Science*, ePub ahead of print.

Dvijotham K, Winkens J, Barsbey M, et al (2023) Enhancing the reliability and accuracy of ai-enabled diagnosis via complementarity-driven deferral to clinicians. *Nature Medicine* 29(7):1814–1820.

FDA (2024) Artificial intelligence and machine learning (AI/ML)-enabled medical devices. https://bit.ly/FDA-AI-ML, published by the U.S. Food and Drug Administration. Accessed: December 11, 2024.

Freeman M, Robinson S, Scholtes S (2021) Gatekeeping, fast and slow: An empirical study of referral errors in the emergency department. *Management Science* 67(7):4209–4232.

Fügener A, Grahl J, Gupta A, Ketter W (2022) Cognitive challenges in human-artificial intelligence collaboration: Investigating the path toward productive delegation. *Information Systems Research* .

Goh E, Gallo R, Hom J, Strong E, Weng Y, Kerman H, Cool JA, Kanjee Z, Parsons AS, Ahuja N, Horvitz E, Yang D, Milstein A, Olson APJ, Rodman A, Chen JH (2024) Large language model influence on diagnostic reasoning: A randomized clinical trial. *JAMA Network Open* 7(10).

Grand-Clèment J, Pauphilet J (2023) The best decisions are not the best advice: Making adherence-aware recommendations. URL http://dx.doi.org/10.48550/ARXIV.2209.01874.

Gurkan H, de Véricourt F (2022) Contracting, pricing, and data collection under the AI flywheel effect. *Management Science* 68(12):8791–8808.

Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57(1):97–109.

Hathaway BA, Kagan E, Dada M (2023) The gatekeeper's dilemma: "When should I transfer this customer?". *Operations Research* 71(3):843–859.

Ho S, Doig GS, Ly A (2022) Attitudes of optometrists towards artificial intelligence for the diagnosis of retinal disease: A cross-sectional mail-out survey. *Ophthalmic and Physiological Optics* 42(6):1170–1179.

Jin L, Tang R, Ye H, Yi J, Zhong S (2023) Path Dependency in Physician Decisions. *The Review of Economic Studies* rdad096.

Kahneman D, Miller DT (1986) Norm theory: Comparing reality to its alternatives. *Psychological Review* 93(2):136–153.

Kovalyk O, Morales-Sánchez J, Verdú-Monedero R, Sellés-Navarro I, Palazón-Cabanes A, Sancho-Gómez JL (2022) PAPILA: Dataset with fundus images and clinical data of both eyes of the same patient for glaucoma assessment. *Scientific Data* 9(1).

Lai J, Xu L, Fang X, Dai T (2024) Regulating adaptive medical artificial intelligence: Can less oversight lead to greater compliance? Working paper, Johns Hopkins University.

Lee HL, Rammohan SV, Sept L (2013) Innovative logistics in extreme conditions: The case of health care delivery in Gambia. Bookbinder JH, ed., *Handbook of Global Logistics: Transportation in International Supply Chains* (Springer New York).

Lee HL, Tang CS (1997) Modelling the costs and benefits of delayed product differentiation. *Management Science* 43(1):40–53.

Lee HL, Tang CS (2018) Socially and environmentally responsible value chain innovations: New operations management research opportunities. *Management Science* 64(3):983–996.

Li Y, Dai T, Qi X (2022) A theory of interior peaks: Activity sequencing and selection for service design. *Manufacturing & Service Operations Management* 24(2):993–1001.

Lieder F, Griffiths TL, Goodman ND (2012) Burn-in, bias, and the rationality of anchoring. *NIPS*, URL https://api.semanticscholar.org/CorpusID:3128634.

Lim JI, Regillo CD, Sadda SR, Ipp E, Bhaskaranand M, Ramachandra C, Solanki K, et al. (2023) Artificial intelligence detection of diabetic retinopathy. *Ophthalmology Science* 3(1):100228.

Luan S, Singh S, Dai T (2024) Algorithmic bias and physician liability. Working paper, Johns Hopkins University.

Ly DP, Shekelle PG, Song Z (2023) Evidence for Anchoring Bias During Physician Decision-Making. *JAMA Internal Medicine* 183(8):818–823.

Mathenge W, Whitestone N, Nkurikiye J, Patnaik JL, Piyasena P, Uwaliraye P, Lanouette G, et al. (2022) Impact of artificial intelligence assessment of diabetic retinopathy on referral service uptake in a low-resource setting. *Ophthalmology Science* 2(4):100168.

Mehta KM, Rerolle F, Rammohan SV, Albohm DC, Muwowo G, Moseson H, Sept L, Lee HL, Bendavid E (2016) Systematic motorcycle management and health care delivery: A field trial. *American Journal of Public Health* 106(1):87–94.

Miao S, Yeung A, Tang CS, Lawson B, Cheng T (2024) Does investment in AI talent pay? An examination of efficiency and spillover effects in manufacturing supply chains. Working paper.

Mullainathan S, Obermeyer Z (2021) Diagnosing physician error: A machine learning approach to low-value health care. *Quarterly Journal of Economics* 137(2):679–727.

Ogdie AR, Reilly JB, Pang WG, Keddem S, Barg FK, Von Feldt JM, Myers JS (2012) Seen through their eyes: Residents' reflections on the cognitive and contextual components of diagnostic errors in medicine. *Academic Medicine* 87(10):1361–1367.

Olsder W, Martagan T, Tang CS (2023) Improving access to rare disease treatments: Subsidy, pricing, and payment schemes. *Management Science* 69(9):5256–5274.

Orfanoudaki A, Saghafian S, Song K, Chakkera HA, Cook C (2022) Algorithm, human, or the centaur: How to enhance clinical care? *HKS Working Paper* .

Price WN, Gerke S, Cohen IG (2019) Potential liability for physicians using artificial intelligence. *JAMA* 322(18):1765–1766.

Rastogi C, Zhang Y, Wei D, Varshney KR, Dhurandhar A, Tomsett R (2022) Deciding fast and slow: The role of cognitive biases in ai-assisted decision-making. *Proc. ACM Hum.-Comput. Interact.* 6(CSCW1).

Ruamviboonsuk P, Krause J, Chotcomwongse P, Sayres R, Raman R, Widner K, Campana BJL, et al. (2019) Deep learning versus human graders for classifying diabetic retinopathy severity in a nationwide screening program. *npj Digital Medicine* 2(1).

Shumsky RA, Pinker EJ (2003) Gatekeepers and referrals in services. *Management Science* 49(7):839–856.

Sodhi MS, Tang CS (2024) Seeking and exploiting synergies among the un sustainability development goals: Research opportunities for operations management. *Production and Operations Management* 33(4):962–978.

Sun H, Saeedi P, Karuranga S, Pinkepank M, Ogurtsova K, Duncan BB, Stein C, et al. (2022) IDF diabetes atlas: Global, regional and country-level diabetes prevalence estimates for 2021 and projections for 2045. *Diabetes Research and Clinical Practice* 183:109119.

Tversky A, Kahneman D (1974) Judgment under uncertainty: Heuristics and biases. *Science* 185(4157):1124–1131.

Wolf RM, Channa R, Abramoff MD, Lehmann HP (2020) Cost-effectiveness of autonomous point-of-care diabetic retinopathy screening for pediatric patients with diabetes. *JAMA Ophthalmology* 138(10):1063.

Xie Y, Nguyen QD, Hamzah H, Lim G, Bellemo V, Gunasekeran DV, Yip MYT, et al. (2020) Artificial intelligence for teleophthalmology-based diabetic retinopathy screening in a national programme: an economic analysis modelling study. *Lancet Digital Health* 2(5):e240–e249.

Appendices

A1: Anchoring and Adjustment in Extreme Cases

A1.1. AI As a Gatekeeper

We analyze the performance of AI as a gatekeeper in two extreme cases, where prior $p=0$ or $p=1$.

Case 1: Let $p=0$. Suppose AI as a gatekeeper gives a positive outcome, with probability $(1-\beta_{AI})$:

1. The probability that the specialist also gives a positive outcome is $(1-\beta_S)$.
2. The probability that the specialist gives a negative outcome is β_S. The probability that the specialist accepts their negative outcome over AI's positive outcome is
$$\min\left\{1, \frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right\},$$
which equals 1 as $p\to 0$.
3. Therefore, the overall probability of a positive outcome as $p\to 0$ is $(1-\beta_{AI})(1-\beta_S)$.

Case 2: Let $p=1$. Suppose AI as a gatekeeper gives a negative outcome, with probability $(1-\alpha_{AI})$. If AI gives a positive outcome,

1. The probability that the specialist also gives a positive outcome is α_S.
2. The probability that the specialist gives a negative outcome is $1-\alpha_S$. The probability that the specialist accepts their negative outcome over AI's positive outcome is
$$\min\left\{1, \frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right\},$$
which equals 0 as $p\to 1$.

Thus, a negative outcome occurs when AI gives a negative outcome with probability $(1-\alpha_{AI})$.

A1.2. AI as a Second Opinion

We now consider AI as a second opinion in the extreme cases of $p=0$ and $p=1$.

Case 1: Let $p=0$. Suppose the specialist gives a positive outcome, with probability $(1-\beta_S)$:

1. The probability that AI also gives a positive outcome is $(1-\beta_{AI})$.
2. The probability that AI gives a negative outcome is β_{AI}. The probability that the specialist accepts AI's negative outcome over their own positive outcome is
$$\min\left\{1, \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})}\right\},$$
which equals 1 as $p\to 0$.

Therefore, the overall probability of a positive outcome as $p\to 0$ is $(1-\beta_{AI})(1-\beta_S)$.

Case 2: Let $p=1$. Suppose the specialist gives a negative outcome, with probability $(1-\alpha_S)$:

1. The probability that AI also gives a negative outcome is $(1-\alpha_{AI})$.
2. The probability that AI gives a positive outcome is α_{AI}. The probability that the specialist accepts AI's positive outcome over their own negative outcome is
$$\min\left\{1, \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\},$$
which equals 1 as $p\to 1$.

3. Therefore, the overall probability of a negative outcome as $p \to 1$ is $(1-\alpha_{AI})(1-\alpha_S)$.

Suppose the specialist gives a positive outcome:

1. The probability that AI gives a negative outcome is $1-\alpha_{AI}$. The probability that the specialist accepts AI's negative outcome over their own positive outcome is
$$\min\left\{1, \frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right\},$$
which equals 0 as $p \to 1$.

True Condition = 1 P(Missed Diagnosis)		
Gatekeeper AI	Second Opinion AI	No AI
$(1-\alpha_{AI})+\alpha_{AI}(1-\alpha_S)\min\left\{1, \frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right\}$	$(1-\alpha_S)(1-\alpha_{AI})+\alpha_S(1-\alpha_{AI})\min\left\{1, \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})}\right\}$ $+(1-\alpha_S)\alpha_{AI}\left(1-\min\left\{1, \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\}\right)$	$1-\alpha_S$

Table A1 Probabilities of Missed Diagnosis

True Condition = 0 P(Unnecessary Treatment)		
Gatekeeper AI	Second Opinion AI	No AI
$(1-\beta_{AI})(1-\beta_S)$ $(1-\beta_{AI})\beta_S\left(1-\min\left\{1, \frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right\}\right)$	$(1-\beta_S)(1-\beta_{AI})+(1-\beta_S)\beta_{AI}\left(1-\min\left\{1, \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})}\right\}\right)$ $+\beta_S(1-\beta_{AI})\min\left\{1, \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\}$	$1-\beta_S$

Table A2 Probabilities of Unnecessary Treatment

A2: Technical Proofs

Proof of Proposition 1: From eqs. (2) and (3), the probabilities of missed diagnoses under gatekeeper AI ($P_M^G(p)$) and under specialist alone ($P_M^0(p)$) equal
$$P_M^G(p) = \begin{cases} p(1-\alpha_{AI}\alpha_S) & if\ p \le \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)} \\ (p(1-\alpha_{AI})+(1-p)\alpha_{AI}(1-\beta_{AI})) & otherwise \end{cases}, \quad P_M^0(p) = p(1-\alpha_S).$$
The expected probability of a missed diagnosis under specialist alone is $p(1-\alpha_S)$. Because $p(1-\alpha_S) < p(1-\alpha_{AI}\alpha_S)$, AI as a gatekeeper reduces the probability of a missed diagnosis if and only if
$$p(1-\alpha_S) > p(1-\alpha_{AI})+(1-p)\alpha_{AI}(1-\beta_{AI})\ and\ p > \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}.$$
Equivalently,
$$p((1-\beta_{AI})+(1-\frac{\alpha_S}{\alpha_{AI}})) > (1-\beta_{AI}).$$
If $(1-\beta_{AI})+(1-\frac{\alpha_S}{\alpha_{AI}}) \le 0$, no feasible value of p exists. Otherwise,
$$p > \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\frac{\alpha_S}{\alpha_{AI}})}\ and\ p > \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}.$$

If $\alpha_{AI} > \alpha_S$,

$$(1-\beta_{AI}) + (1 - \frac{\alpha_S}{\alpha_{AI}}) > 0 \textit{ and } 1 > \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\frac{\alpha_S}{\alpha_{AI}})} > \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)}$$

and

$$\tau_M^G = \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\frac{\alpha_S}{\alpha_{AI}})}.$$

If $\alpha_{AI} < \alpha_S$ *and* $(1-\beta_{AI}) + (1-\frac{\alpha_S}{\alpha_{AI}}) > 0$,

$$\frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\frac{\alpha_S}{\alpha_{AI}})} > 1 > \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)};$$

that is, no feasible value of p exists. *Q.E.D.*

Proof of Proposition 2: From eqs. (4) and (5), the probabilities of unnecessary treatment under gatekeeper AI ($P_T^G(p)$) and specialist alone ($P_T^0(p)$) are:

$$P_T^G(p) = (1-p)\cdot\begin{cases}(1-\beta_{AI})(1-\beta_S) & \text{if } p \le \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}\\ (1-\beta_{AI}) - \frac{(1-p)(1-\beta_{AI})^2\beta_S}{p(1-\alpha_S)} & \text{if } p > \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}\end{cases},\ P_T^0(p) = (1-p)(1-\beta_S).$$

Further,

$$P_T^G(p) - P_T^0(p) = (1-p)\cdot\begin{cases}-\beta_{AI}(1-\beta_S) & \text{if } p \le \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}\\ (\beta_S-\beta_{AI}) - \frac{(1-p)(1-\beta_{AI})^2\beta_S}{p(1-\alpha_S)} & \text{if } p > \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}.\end{cases}$$

The difference is always negative or zero if $\beta_{AI} \ge \beta_S$. Otherwise, if $\beta_{AI} < \beta_S$, the difference is positive if

$$(\beta_S - \beta_{AI}) - \frac{(1-p)(1-\beta_{AI})^2\beta_S}{p(1-\alpha_S)} > 0 \textit{ and } p > \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)}.$$

Equivalently,

$$p > \frac{(1-\beta_{AI})^2}{(1-\beta_{AI})^2 + (1-\alpha_S)(1-\frac{\beta_{AI}}{\beta_S})} \textit{ and } p > \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)}.$$

Further,

$$\frac{(1-\beta_{AI})^2}{(1-\beta_{AI})^2 + (1-\alpha_S)(1-\frac{\beta_{AI}}{\beta_S})} > \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)} \iff (1-\alpha_S)(\frac{\beta_{AI}}{\beta_S} - \beta_{AI}) > 0.$$

Therefore,

$$\tau_T^G = \frac{(1-\beta_{AI})^2}{(1-\beta_{AI})^2 + (1-\alpha_S)(1-\frac{\beta_{AI}}{\beta_S})}.$$

Further,

$$\frac{\partial \tau_T^G}{\partial \beta_{AI}} = \frac{(1-\beta_{AI})(\alpha_S)(\beta_{AI} + 1 - 2\beta_S)}{\beta_S(1-\beta_{AI})^2 + (1-\alpha_S)(1-\frac{\beta_{AI}}{\beta_S})^2}.$$

Q.E.D.

Proof of Proposition 3: Simplifying $P_M^S(p)$ eqs. (7) and (8), we obtain the following:

1. If $\frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} \leq \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}$,

$$P_M^S(p) - P_M^0(p) = \begin{cases} [p\alpha_S(1-\alpha_{AI}) - \frac{p^2\alpha_{AI}(1-\alpha_S)^2}{(1-p)(1-\beta_{AI})}] & if\ p < \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})}, \\ p(\alpha_S - \alpha_{AI}) & if\ \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} \leq p \leq \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}, \\ [-p\alpha_{AI}(1-\alpha_S) + (1-p)\alpha_S(1-\beta_S)] & if\ p > \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}. \end{cases}$$

Now,

$$p\alpha_S(1-\alpha_{AI}) - \frac{p^2\alpha_{AI}(1-\alpha_S)^2}{(1-p)(1-\beta_{AI})} \leq 0 \iff p \geq \frac{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)^2 + \alpha_S(1-\alpha_{AI})(1-\beta_{AI})},$$

where

$$\frac{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)^2 + \alpha_S(1-\alpha_{AI})(1-\beta_{AI})} \leq \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} \iff \alpha_S - \alpha_{AI} \leq 0.$$

Further,

$$-p\alpha_{AI}(1-\alpha_S) + (1-p)\alpha_S(1-\beta_S) \leq 0 \iff p \geq \frac{\alpha_S(1-\beta_S)}{\alpha_S(1-\beta_S) + \alpha_{AI}(1-\alpha_S)},$$

where

$$\frac{\alpha_S(1-\beta_S)}{\alpha_S(1-\beta_S) + \alpha_{AI}(1-\alpha_S)} \geq \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} \iff \alpha_S - \alpha_{AI} \geq 0.$$

Therefore, we conclude the following

- If $\alpha_S - \alpha_{AI} \leq 0$,

$$P_M^S(p) - P_M^0(p) \leq 0 \iff p \geq \frac{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)^2 + \alpha_S(1-\alpha_{AI})(1-\beta_{AI})}.$$

- If $\alpha_S - \alpha_{AI} \geq 0$,

$$P_M^S(p) - P_M^0(p) \leq 0 \iff p \geq \frac{\alpha_S(1-\beta_S)}{\alpha_S(1-\beta_S) + \alpha_{AI}(1-\alpha_S)}.$$

2. Alternatively, if $\frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} > \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}$,

$$P_M^S(p) - P_M^0(p) = \begin{cases} [p\alpha_S(1-\alpha_{AI}) - \frac{p^2\alpha_{AI}(1-\alpha_S)^2}{(1-p)(1-\beta_{AI})}] & if\ p \leq \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}, \\ [-\frac{p^2\alpha_{AI}(1-\alpha_S)^2}{(1-p)(1-\beta_{AI})} + (1-p)\alpha_S(1-\beta_S)] & if\ \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} < p < \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})}, \\ [-p\alpha_{AI}(1-\alpha_S) + (1-p)\alpha_S(1-\beta_S)] & if\ p \geq \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})}. \end{cases}$$

Now,

$$p\alpha_S(1-\alpha_{AI}) - \frac{p^2\alpha_{AI}(1-\alpha_S)^2}{(1-p)(1-\beta_{AI})} \leq 0 \iff p \geq \frac{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)^2 + \alpha_S(1-\alpha_{AI})(1-\beta_{AI})},$$

where

$$\frac{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)^2 + \alpha_S(1-\alpha_{AI})(1-\beta_{AI})} \leq \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} \iff$$

$$\alpha_S(1-\alpha_{AI})^2(1-\beta_{AI}) \leq \alpha_{AI}(1-\alpha_S)^2(1-\beta_S).$$

Similarly,

$$-\frac{p^2\alpha_{AI}(1-\alpha_S)^2}{(1-p)(1-\beta_{AI})} + (1-p)\alpha_S(1-\beta_S) \leq 0 \iff p \geq \frac{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}}{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})} + \sqrt{\alpha_{AI}(1-\alpha_S)^2}},$$

where

$$\frac{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}}{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}+\sqrt{\alpha_{AI}(1-\alpha_S)^2}} \geq \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} \iff$$

$$\alpha_S(1-\alpha_{AI})^2(1-\beta_{AI}) \geq \alpha_{AI}(1-\alpha_S)^2(1-\beta_S)$$

and

$$\frac{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}}{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}+\sqrt{\alpha_{AI}(1-\alpha_S)^2}} \leq \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} \iff \alpha_S(1-\beta_S) \leq \alpha_{AI}(1-\beta_{AI}).$$

Finally,

$$-p\alpha_{AI}(1-\alpha_S)+(1-p)\alpha_S(1-\beta_S) \leq 0 \iff p \geq \frac{\alpha_S(1-\beta_S)}{\alpha_S(1-\beta_S)+\alpha_{AI}(1-\alpha_S)},$$

where

$$\frac{\alpha_S(1-\beta_S)}{\alpha_S(1-\beta_S)+\alpha_{AI}(1-\alpha_S)} \geq \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} \iff \alpha_S(1-\beta_S) \geq \alpha_{AI}(1-\beta_{AI}).$$

Now, $\frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} > \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} \iff (1-\alpha_{AI})(1-\beta_{AI}) > (1-\alpha_S)(1-\beta_S)$.

- If $\alpha_S(1-\alpha_{AI})^2(1-\beta_{AI}) \leq \alpha_{AI}(1-\alpha_S)^2(1-\beta_S)$,

$$\alpha_S(1-\alpha_{AI}) < \alpha_{AI}(1-\alpha_S) \implies \alpha_S < \alpha_{AI}$$

and

$$\frac{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}}{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}+\sqrt{\alpha_{AI}(1-\alpha_S)^2}} \leq \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} \leq \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} \implies$$

$$\beta_S - \beta_{AI} \geq 0$$

and

$$\alpha_S(1-\beta_S) < \alpha_{AI}(1-\beta_{AI}).$$

Therefore, if $\alpha_S(1-\alpha_{AI})^2(1-\beta_{AI}) \leq \alpha_{AI}(1-\alpha_S)^2(1-\beta_S)$,

$$P_M^S(p) - P_M^0(p) \leq 0 \iff p \geq \frac{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)^2+\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}.$$

- If $\alpha_S(1-\alpha_{AI})^2(1-\beta_{AI}) > \alpha_{AI}(1-\alpha_S)^2(1-\beta_S)$,

$$\frac{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}{\alpha_{AI}(1-\alpha_S)^2+\alpha_S(1-\alpha_{AI})(1-\beta_{AI})} > \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}$$

and

$$\frac{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}}{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}+\sqrt{\alpha_{AI}(1-\alpha_S)^2}} \geq \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)}.$$

Further, if $\beta_S - \beta_{AI} < 0$,

$$(1-\alpha_{AI})(1-\beta_{AI}) > (1-\alpha_S)(1-\beta_S) \implies \alpha_S > \alpha_{AI} \implies \alpha_S(1-\beta_S) > \alpha_{AI}(1-\beta_S).$$

Therefore, if $\alpha_S(1-\alpha_{AI})^2(1-\beta_{AI}) > \alpha_{AI}(1-\alpha_S)^2(1-\beta_S)$ and $\beta_S < \beta_{AI}$,

$$P_M^S(p) - P_M^0(p) \leq 0 \iff p \geq \frac{\alpha_S(1-\beta_S)}{\alpha_S(1-\beta_S)+\alpha_{AI}(1-\alpha_S)}.$$

Alternately, if $\alpha_S(1-\alpha_{AI})^2(1-\beta_{AI}) > \alpha_{AI}(1-\alpha_S)^2(1-\beta_S)$, $\beta_S - \beta_{AI} \geq 0$ and $\alpha_S(1-\beta_S) < \alpha_{AI}(1-\beta_{AI})$,

$$P_M^S(p) - P_M^0(p) \leq 0 \iff p \geq \frac{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}}{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}+\sqrt{\alpha_{AI}(1-\alpha_S)^2}}.$$

Finally, if $\alpha_S(1-\alpha_{AI})^2(1-\beta_{AI}) > \alpha_{AI}(1-\alpha_S)^2(1-\beta_S)$, $\beta_S - \beta_{AI} \geq 0$ and $\alpha_S(1-\beta_S) \geq \alpha_{AI}(1-\beta_{AI})$,

$$P_M^S(p) - P_M^0(p) \leq 0 \iff$$

$$p \in [\frac{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}}{\sqrt{\alpha_S(1-\beta_S)(1-\beta_{AI})}+\sqrt{\alpha_{AI}(1-\alpha_S)^2}}, \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})}] \cup [\frac{\alpha_S(1-\beta_S)}{\alpha_S(1-\beta_S)+\alpha_{AI}(1-\alpha_S)}, 1]$$

Q.E.D.

	Condition	Priors where AI as a second opinion reduces (or maintains) the probability of unnecessary treatment
1	$(1-\alpha_{AI})(1-\beta_{AI}) \leq (1-\alpha_S)(1-\beta_S)$, $\beta_S > \beta_{AI}$	$[0, \frac{\beta_{AI}(1-\beta_S)}{\beta_{AI}(1-\beta_S)+\beta_S(1-\alpha_S)}]$
2	$(1-\alpha_{AI})(1-\beta_{AI}) \leq (1-\alpha_S)(1-\beta_S)$, $\beta_S \leq \beta_{AI}$	$[0, \frac{(1-\beta_S)^2\beta_{AI}}{(1-\beta_S)^2\beta_{AI}+\beta_S(1-\beta_{AI})(1-\alpha_{AI})}]$
3	$(1-\alpha_{AI})(1-\beta_{AI}) > (1-\alpha_S)(1-\beta_S)$, $\beta_{AI}(1-\alpha_{AI}) \leq \beta_S(1-\alpha_S)$, $(1-\alpha_{AI})(1-\beta_{AI})^2\beta_S > (1-\alpha_S)(1-\beta_S)^2\beta_{AI}$	$[0, \frac{(1-\beta_S)\beta_{AI}}{(1-\beta_S)\beta_{AI}+(1-\alpha_S)\beta_S}]$
4	$(1-\alpha_{AI})(1-\beta_{AI}) > (1-\alpha_S)(1-\beta_S)$, $\beta_{AI}(1-\alpha_{AI}) > \beta_S(1-\alpha_S)$, $(1-\alpha_{AI})(1-\beta_{AI})^2\beta_S \leq (1-\alpha_S)(1-\beta_S)^2\beta_{AI}$	$[0, \frac{(1-\beta_S)^2\beta_{AI}}{(1-\beta_S)^2\beta_{AI}+\beta_S(1-\beta_{AI})(1-\alpha_{AI})}]$
5	$(1-\alpha_{AI})(1-\beta_{AI}) > (1-\alpha_S)(1-\beta_S)$, $\beta_{AI}(1-\alpha_{AI}) > \beta_S(1-\alpha_S)$, $(1-\alpha_{AI})(1-\beta_{AI})^2\beta_S > (1-\alpha_S)(1-\beta_S)^2\beta_{AI}$	$[0, \frac{\sqrt{(1-\beta_S)^2\beta_{AI}}}{\sqrt{(1-\beta_S)^2\beta_{AI}}+\sqrt{(1-\alpha_{AI})(1-\alpha_S)\beta_S}}]$

Table A3 Proposition 4

Proof of Proposition 4: Because

$$\min\left\{1, \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})}\right\} = \begin{cases} 1 & \text{if } p \leq \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} \\ \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})} & \text{if } p > \frac{1-\beta_S}{(1-\alpha_{AI})+(1-\beta_S)} \end{cases}$$

and,

$$min\left\{1, \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\} = \begin{cases} \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})} & \text{if } p \leq \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})} \\ 1 & \text{if } p > \frac{1-\beta_{AI}}{(1-\alpha_S)+(1-\beta_{AI})}, \end{cases}$$

- If $\frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} \leq \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}}$,

$$P_T^S(p) = \begin{cases} (1-p)(1-\beta_S)(1-\beta_{AI}) + p(1-\alpha_S)\beta_S & if\ p \leq \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} \\ (1-p)(1-\beta_{AI}) & if\ \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} < p < \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} \\ (1-p)(1-\beta_{AI}) + (1-p)(1-\beta_S)\beta_{AI} - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} & if\ p \geq \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} \end{cases}$$

and

$$P_T^S(p) - P_T^0(p) = \begin{cases} -(1-p)(1-\beta_S)\beta_{AI} + p(1-\alpha_S)\beta_S & if\ p \leq \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} \\ (1-p)(\beta_S - \beta_{AI}) & if\ \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} < p < \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} \\ (1-p)(\beta_S - \beta_{AI}) + (1-p)(1-\beta_S)\beta_{AI} - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} & if\ p \geq \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}}. \end{cases}$$

Now,

$$-(1-p)(1-\beta_S)\beta_{AI} + p(1-\alpha_S)\beta_S \leq 0 \iff p \leq \frac{(1-\beta_S)\beta_{AI}}{(1-\beta_S)\beta_{AI} + (1-\alpha_S)\beta_S},$$

where

$$\frac{(1-\beta_S)\beta_{AI}}{(1-\beta_S)\beta_{AI} + (1-\alpha_S)\beta_S} \leq \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} \iff \beta_S \geq \beta_{AI}.$$

Further,

$$(1-p)(\beta_S - \beta_{AI}) \leq 0 \iff \beta_S \leq \beta_{AI}.$$

Finally,

$$(1-p)(\beta_S - \beta_{AI}) + (1-p)(1-\beta_S)\beta_{AI} - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} \leq 0 \iff p \leq \frac{(1-\beta_S)^2\beta_{AI}}{(1-\beta_S)^2\beta_{AI} + \beta_S(1-\beta_{AI})(1-\alpha_{AI})},$$

where,

$$\frac{(1-\beta_S)^2\beta_{AI}}{(1-\beta_S)^2\beta_{AI} + \beta_S(1-\beta_{AI})(1-\alpha_{AI})} \geq \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} \iff \beta_S \leq \beta_{AI}.$$

Therefore, if $\beta_{AI} < \beta_S$,

$$P_T^S(p) - P_T^0(p) \leq 0 \iff p \leq \frac{(1-\beta_S)\beta_{AI}}{(1-\beta_S)\beta_{AI} + (1-\alpha_S)\beta_S}.$$

If $\beta_{AI} > \beta_S$,

$$P_T^S(p) - P_T^0(p) \leq 0 \iff p \leq \frac{(1-\beta_S)^2\beta_{AI}}{(1-\beta_S)^2\beta_{AI} + \beta_S(1-\beta_{AI})(1-\alpha_{AI})}.$$

If $\beta_{AI} = \beta_S$,

$$P_T^S(p) - P_T^0(p) \leq \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}}.$$

- Alternately, if $\frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} > \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}}$,

$$P_T^S(p) = \begin{cases} (1-p)(1-\beta_S)(1-\beta_{AI}) + p(1-\alpha_S)\beta_S & if\ p \leq \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} \\ (1-p)(1-\beta_S) + p(1-\alpha_S)\beta_S - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} & if\ \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} < p < \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} \\ (1-p)(1-\beta_{AI}) + (1-p)(1-\beta_S)\beta_{AI} - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} & if\ p \geq \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} \end{cases}$$

and

$$P_T^S(p) - P_T^0(p) = \begin{cases} -(1-p)(1-\beta_S)\beta_{AI} + p(1-\alpha_S)\beta_S & if\ p \leq \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} \\ p(1-\alpha_S)\beta_S - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} & if\ \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} < p < \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} \\ (1-p)(\beta_S-\beta_{AI}) + (1-p)(1-\beta_S)\beta_{AI} - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} & if\ p \geq \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S}. \end{cases}$$

Now,

$$-(1-p)(1-\beta_S)\beta_{AI} + p(1-\alpha_S)\beta_S \leq 0 \iff p \leq \frac{(1-\beta_S)\beta_{AI}}{(1-\beta_S)\beta_{AI} + (1-\alpha_S)\beta_S},$$

where

$$\frac{(1-\beta_S)\beta_{AI}}{(1-\beta_S)\beta_{AI} + (1-\alpha_S)\beta_S} \leq \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} \iff \beta_{AI}(1-\alpha_{AI}) \leq \beta_S(1-\alpha_S).$$

Further,

$$p(1-\alpha_S)\beta_S - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} \leq 0 \iff p \leq \frac{\sqrt{(1-\beta_S)^2\beta_{AI}}}{\sqrt{(1-\beta_S)^2\beta_{AI}} + \sqrt{(1-\alpha_{AI})(1-\alpha_S)\beta_S}},$$

where

$$\frac{\sqrt{(1-\beta_S)^2\beta_{AI}}}{\sqrt{(1-\beta_S)^2\beta_{AI}} + \sqrt{(1-\alpha_{AI})(1-\alpha_S)\beta_S}} > \frac{1-\beta_S}{1-\beta_S+1-\alpha_{AI}} \iff \beta_{AI}(1-\alpha_{AI}) > \beta_S(1-\alpha_S)$$

and

$$\frac{\sqrt{(1-\beta_S)^2\beta_{AI}}}{\sqrt{(1-\beta_S)^2\beta_{AI}} + \sqrt{(1-\alpha_{AI})(1-\alpha_S)\beta_S}} < \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} \iff (1-\alpha_{AI})(1-\beta_{AI})^2\beta_S > (1-\alpha_S)(1-\beta_S)^2\beta_{AI}.$$

Finally,

$$(1-p)(\beta_S-\beta_{AI}) + (1-p)(1-\beta_S)\beta_{AI} - \frac{(1-p)^2(1-\beta_S)^2\beta_{AI}}{p(1-\alpha_{AI})} \leq 0 \iff p \leq \frac{(1-\beta_S)^2\beta_{AI}}{(1-\beta_S)^2\beta_{AI} + \beta_S(1-\beta_{AI})(1-\alpha_{AI})},$$

where

$$\frac{(1-\beta_S)^2\beta_{AI}}{(1-\beta_S)^2\beta_{AI} + \beta_S(1-\beta_{AI})(1-\alpha_{AI})} \geq \frac{1-\beta_{AI}}{1-\beta_{AI}+1-\alpha_S} \iff (1-\alpha_{AI})(1-\beta_{AI})^2\beta_S \leq (1-\alpha_S)(1-\beta_S)^2\beta_{AI}.$$

Proof of Proposition 5: The difference in costs $(C_G(p) - C_0(p))$ is given by

$$C_G(p) - C_0(p) = \begin{cases} C_M p(1-\alpha_{AI})\alpha_S - C_T(1-p)\beta_{AI}(1-\beta_S), & \text{if } p \leq \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}, \\ C_M p(\alpha_S - \alpha_{AI}) + C_M(1-p)\alpha_{AI}(1-\beta_{AI}) \\ -C_T(1-p)\beta_{AI}(1-\beta_S) + C_T(1-p)(1-\beta_{AI})\beta_S\left(1 - \frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}\right), & \text{if } p > \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}. \end{cases}$$

For $p \leq \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}$,

$$C_G(p) - C_0(p) \leq 0 \iff \begin{cases} p \leq \frac{C_T\beta_{AI}(1-\beta_S)}{C_M(1-\alpha_{AI})\alpha_S + C_T\beta_{AI}(1-\beta_S)} & \text{if } \frac{C_M}{C_T} \geq \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})} \\ p \leq \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)} & \text{if } \frac{C_M}{C_T} < \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}. \end{cases} \tag{A1}$$

For $p > \frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}$,

$$C_G(p) - C_0(p) \leq 0 \iff \frac{p^2}{(1-p)^2}C_M(\alpha_S-\alpha_{AI})(1-\alpha_S) + \frac{p}{1-p}(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI}) - C_T(\beta_{AI}-\beta_S)) \\ - C_T\beta_S(1-\beta_{AI})^2 \leq 0.$$

Condition	Priors where AI as a gatekeeper reduces (or maintains) the total expected cost of a missed diagnosis and unnecessary treatment
$\alpha_S - \alpha_{AI} > 0$, $C_M\alpha_S(1-\alpha_{AI})(1-\beta_{AI}) - C_T\beta_{AI}(1-\alpha_S)(1-\beta_S) \leq 0$	$[0, \frac{\sqrt{D}-(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))}{\sqrt{D}-(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))+2C_M(\alpha_S-\alpha_{AI})(1-\alpha_S)}]$
$\alpha_S - \alpha_{AI} > 0$, $C_M\alpha_S(1-\alpha_{AI})(1-\beta_{AI}) - C_T\beta_{AI}(1-\alpha_S)(1-\beta_S) > 0$	$[0, \frac{C_T\beta_{AI}(1-\beta_S)}{C_M(1-\alpha_{AI})\alpha_S+C_T\beta_{AI}(1-\beta_S)}]$
$\alpha_S - \alpha_{AI} < 0$, $D < 0$ $C_M\alpha_S(1-\alpha_{AI})(1-\beta_{AI}) - C_T\beta_{AI}(1-\alpha_S)(1-\beta_S) < 0$	$[0, 1]$
$\alpha_S - \alpha_{AI} \leq 0$, $D \geq 0$ $C_M\alpha_S(1-\alpha_{AI})(1-\beta_{AI}) - C_T\beta_{AI}(1-\alpha_S)(1-\beta_S) > 0$	$[0, \frac{C_T\beta_{AI}(1-\beta_S)}{C_M(1-\alpha_{AI})\alpha_S+C_T\beta_{AI}(1-\beta_S)}]\cup$ $[\frac{\sqrt{D}+(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))}{\sqrt{D}+(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))+2C_M(\alpha_{AI}-\alpha_S)(1-\alpha_S)}, 1]$
$\alpha_S - \alpha_{AI} < 0$, $D \geq 0$ $C_M\alpha_S(1-\alpha_{AI})(1-\beta_{AI}) - C_T\beta_{AI}(1-\alpha_S)(1-\beta_S) \leq 0$ $C_M(1-\beta_{AI})(\alpha_{AI}-\alpha_S) - C_M\alpha_S(1-\alpha_{AI})(1-\beta_{AI})$ $-C_T(\beta_S-\beta_{AI})(1-\alpha_S) \geq 0$	$[0, 1]$
$\alpha_S - \alpha_{AI} \leq 0$, $D \geq 0$ $C_M\alpha_S(1-\alpha_{AI})(1-\beta_{AI}) - C_T\beta_{AI}(1-\alpha_S)(1-\beta_S) \leq 0$ $C_M(1-\beta_{AI})(\alpha_{AI}-\alpha_S) - C_M\alpha_S(1-\alpha_{AI})(1-\beta_{AI})$ $-C_T(\beta_S-\beta_{AI})(1-\alpha_S) < 0$	$[0, \frac{\sqrt{D}-(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))}{\sqrt{D}-(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))+2C_M(\alpha_S-\alpha_{AI})(1-\alpha_S)}]$ $\cup[\frac{\sqrt{D}+(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))}{\sqrt{D}+(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))+2C_M(\alpha_{AI}-\alpha_S)(1-\alpha_S)}, 1]$

Table A4 Gatekeeper vs. No AI for $\beta_S > \beta_{AI}$,
$D = (1-\alpha_S)^2(C_M\alpha_{AI}(1-\beta_{AI}) + C_T(\beta_S - \beta_{AI}))^2 + 4C_M C_T(\alpha_S - \alpha_{AI})(1-\alpha_S)\beta_S(1-\beta_{AI})^2$; $\lim_{\alpha_S \to \alpha_{AI}} D > 0$ **and**

$$\begin{aligned}
&\lim_{\alpha_S \to \alpha_{AI}} \frac{\sqrt{D} - (1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI}) + C_T(\beta_S - \beta_{AI}))}{\sqrt{D} - (1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI}) + C_T(\beta_S - \beta_{AI})) + 2C_M(\alpha_S - \alpha_{AI})(1-\alpha_S)} \\
&= \frac{C_T\beta_S(1-\beta_{AI})^2}{(C_M(1-\alpha_S)\alpha_{AI}(1-\beta_{AI}) + C_T\beta_S(1-\beta_{AI})^2 + C_T(1-\alpha_S)(\beta_S - \beta_{AI}))} \\
&\lim_{\alpha_S \to \alpha_{AI}} \frac{\sqrt{D} + (1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI}) + C_T(\beta_S - \beta_{AI}))}{\sqrt{D} + (1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI}) + C_T(\beta_S - \beta_{AI})) + 2C_M(\alpha_{AI} - \alpha_S)(1-\alpha_S)} = 1.
\end{aligned}$$

Letting $\frac{p}{1-p} = X \iff p = \frac{X}{X+1}$,

$$p > \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)} \;\&\; C_G(p) - C_0(p) \leq 0 \iff f(X) = C_M(\alpha_S - \alpha_{AI})(1-\alpha_S)X^2 + (1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI}) - C_T(\beta_{AI} -$$

$$- C_T\beta_S(1-\beta_{AI})^2 \leq 0 \;\&\; X > \frac{(1-\beta_{AI})}{(1-\alpha_S)}.$$

If $\alpha_S = \alpha_{AI}$ and $(C_M\alpha_{AI}(1-\beta_{AI}) - C_T(\beta_{AI} - \beta_S)) > 0$ (which is always true when $\beta_S \geq \beta_{AI}$), then $f(X) \leq 0$ if

$$X \leq \frac{C_T\beta_S(1-\beta_{AI})^2}{(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI}) - C_T(\beta_{AI} - \beta_S))} \;\&\; X > \frac{(1-\beta_{AI})}{(1-\alpha_S)}.$$

Further, if

$$\frac{C_T\beta_S(1-\beta_{AI})^2}{(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})-C_T(\beta_{AI}-\beta_S))} \geq \frac{(1-\beta_{AI})}{(1-\alpha_S)} \iff C_M\alpha_{AI}(1-\beta_{AI})-C_T\beta_{AI}(1-\beta_S)\leq 0,$$

then

$$C_G(p)-C_0(p)\leq 0 \iff p\in\left[0,\min\left\{\frac{C_T\beta_{AI}(1-\beta_S)}{C_M(1-\alpha_{AI})\alpha_S+C_T\beta_{AI}(1-\beta_S)},\frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)}\right\}\right] \cup\left[\frac{(1-\beta_{AI})}{(1-\beta_{AI})+(1-\alpha_S)},\frac{C_T\beta_S(1-\beta_{AI})^2}{C_M(1-\alpha_S)\alpha_{AI}(1-\beta_{AI})+C_T\beta_S(1-\beta_{AI})^2-C_T(1-\alpha_S)(\beta_{AI}-\beta_S)}\right].$$

Recall that

$$f(X)=C_M(\alpha_S-\alpha_{AI})(1-\alpha_S)X^2+(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))X-C_T\beta_S(1-\beta_{AI})^2$$

and

$$f\left(\frac{(1-\beta_{AI})}{(1-\alpha_S)}\right)=\frac{C_T\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}{(1-\alpha_S)}\left[\frac{C_M}{C_T}-\frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}\right].$$

When $\alpha_S\neq\alpha_{AI}$, $f(X)=0$ is a quadratic equation with roots

$$X_1^*=\frac{\sqrt{D}-(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))}{2C_M(\alpha_S-\alpha_{AI})(1-\alpha_S)}$$
$$X_2^*=\frac{-\sqrt{D}-(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))}{2C_M(\alpha_S-\alpha_{AI})(1-\alpha_S)},$$

and vertex

$$X^*=\frac{-(1-\alpha_S)(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))}{2C_M(\alpha_S-\alpha_{AI})(1-\alpha_S)},$$

where

$$D=(1-\alpha_S)^2(C_M\alpha_{AI}(1-\beta_{AI})+C_T(\beta_S-\beta_{AI}))^2+4C_MC_T(\alpha_S-\alpha_{AI})(1-\alpha_S)\beta_S(1-\beta_{AI})^2.$$

If $\alpha_S>\alpha_{AI}$, then $f(X)$ is an upward parabola with discriminant $D>0$, $f(0)<0$, and $X_1^*>0$, $X_2^*<0$, with a vertex less than 0.

- *Case 1:* If $\frac{C_M}{C_T}<\frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$, then $f\left(\frac{(1-\beta_{AI})}{(1-\alpha_S)}\right)<0$, and $\frac{(1-\beta_{AI})}{(1-\alpha_S)}<X_1^*$. Therefore, from eq. (A1),

$$C_G(p)-C_0(p)\leq 0 \iff p\in\left[0,\frac{X_1^*}{X_1^*+1}\right].$$

- *Case 2:* If $\frac{C_M}{C_T}\geq\frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$, then $f\left(\frac{(1-\beta_{AI})}{(1-\alpha_S)}\right)\geq 0$, and $\frac{(1-\beta_{AI})}{(1-\alpha_S)}\geq X_1^*$. Therefore, from eq. (A1),

$$C_G(p)-C_0(p)\leq 0 \iff p\in\left[0,\frac{C_T\beta_{AI}(1-\beta_S)}{C_M(1-\alpha_{AI})\alpha_S+C_T\beta_{AI}(1-\beta_S)}\right].$$

If $\alpha_S<\alpha_{AI}$, then $f(X)$ is a downward parabola, with $f(0)<0$ and vertex greater than 0.

- *Case 1:* If $D<0$, then $f(X)<0$ for all $X\geq 0$, implying that $f\left(\frac{(1-\beta_{AI})}{(1-\alpha_S)}\right)<0$, which is equivalent to $\frac{C_M}{C_T}<\frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$. Therefore, from eq. (A1),

$$C_G(p)-C_0(p)\leq 0 \iff p\in[0,1].$$

One example where this case holds is $C_M=C_T=50$, $\alpha_{AI}=0.95$, $\beta_{AI}=0.81$, $\alpha_S=0.75$, $\beta_S=0.95$, $D=-1.097$.

- *Case 2:* If $D \geq 0$, then $0 \leq X_1^* \leq X_2^*$

—If $\frac{C_M}{C_T} \geq \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$, then $f\left(\frac{(1-\beta_{AI})}{(1-\alpha_S)}\right) \geq 0$. Therefore, from eq. (A1),

$$C_G(p) - C_0(p) \leq 0 \iff p \in \left[0, \frac{C_T\beta_{AI}(1-\beta_S)}{C_M(1-\alpha_{AI})\alpha_S + C_T\beta_{AI}(1-\beta_S)}\right] \cup \left[\frac{X_2^*}{X_2^*+1}, 1\right].$$

Note $\alpha_S < \alpha_{AI}$ (a downward parabola) and $f\left(\frac{(1-\beta_{AI})}{(1-\alpha_S)}\right) \geq 0$ together imply $D \geq 0$.

—If $\frac{C_M}{C_T} < \frac{\beta_{AI}(1-\alpha_S)(1-\beta_S)}{\alpha_S(1-\alpha_{AI})(1-\beta_{AI})}$, then $f\left(\frac{(1-\beta_{AI})}{(1-\alpha_S)}\right) < 0$. Further, if $\frac{(1-\beta_{AI})}{(1-\alpha_S)} < X^*$, that is, $C_M(\alpha_{AI} + \alpha_{AI}\alpha_S - 2\alpha_S) < C_T(\beta_S - \beta_{AI})(1-\alpha_S)$, then from eq. (A1),

$$C_G(p) - C_0(p) \leq 0 \iff p \in \left[0, \frac{X_1^*}{X_1^*+1}\right] \cup \left[\frac{X_2^*}{X_2^*+1}, 1\right].$$

However, if $\frac{(1-\beta_{AI})}{(1-\alpha_S)} > X^*$, that is, $C_M(\alpha_{AI} + \alpha_{AI}\alpha_S - 2\alpha_S) > C_T(\beta_S - \beta_{AI})(1-\alpha_S)$, then from eq. (A1),

$$C_G(p) - C_0(p) \leq 0 \iff p \in [0, 1].$$

Q.E.D.

Proof of Lemma 1: If $C_T = 0$, then $C_G(p) - C_S(p) = C_M(P_M^G(p) - P_M^S(p))$. Using eqs. (3) and (7), we have

$$\begin{aligned} C_G(p) - C_S(p) = C_M p\Big[&(\alpha_S - \alpha_{AI}) \\ &+ \alpha_{AI}(1-\alpha_S)\Big(1 + \min\left\{\frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}, \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\}\Big) \\ &- \alpha_S(1-\alpha_{AI})\min\left\{1, \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})}\right\}\Big] \\ = C_M p\Big[&\alpha_S(1-\alpha_{AI})(1 - \min\left\{1, \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})}\right\}) \\ &+ \alpha_{AI}(1-\alpha_S)\min\left\{\frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}, \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\}\Big] \geq 0. \end{aligned}$$

Q.E.D.

Proof of Lemma 2: If $C_M = 0$, then $C_G(p) - C_S(p) = C_T(P_T^G(p) - P_T^S(p))$. Using eqs. (5) and (9), we have

$$\begin{aligned} C_G(p) - C_S(p) = -C_T(1-p)\Big[&\beta_S(1-\beta_{AI})(1 - \min\left\{1, \frac{(1-p)(1-\beta_S)}{p(1-\alpha_{AI})}\right\}) \\ &+ \beta_{AI}(1-\beta_S)\min\left\{\frac{(1-p)(1-\beta_{AI})}{p(1-\alpha_S)}, \frac{p(1-\alpha_S)}{(1-p)(1-\beta_{AI})}\right\}\Big] \leq 0. \end{aligned}$$

Q.E.D.

Proof of Proposition 6: If

$$p_1 = \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)} < \frac{(1-\beta_S)}{(1-\beta_S) + (1-\alpha_{AI})} = p_2,$$

the difference in expected costs of missed diagnoses and unnecessary treatment under AI as a second opinion and AI as a gatekeeper, $C_S(p) - C_G(p)$, is given by

Condition	Priors where AI as a second opinion has total expected cost less than or equal to AI as a gatekeeper
$C_M\alpha_{AI}(1-\alpha_S)(1-\beta_S)$ $-C_T\beta_S(1-\alpha_{AI})(1-\beta_{AI}) \leq 0$	$[\frac{(1-\alpha_{AI})(1-\alpha_S)\left[(C_T\beta_{AI}+C_M\alpha_S)(1-\beta_S)-C_M\alpha_{AI}(1-\beta_{AI})\right]+\sqrt{D'}}{(1-\alpha_{AI})(1-\alpha_S)\left[(C_T\beta_{AI}+C_M\alpha_S)(1-\beta_S)-C_M\alpha_{AI}(1-\beta_{AI})\right]+\sqrt{D'}+2C_M\alpha_S(1-\alpha_S)(1-\alpha_{AI})^2}, 1]$
$C_M\alpha_{AI}(1-\alpha_S)(1-\beta_S)$ $-C_T\beta_S(1-\alpha_{AI})(1-\beta_{AI}) > 0$	$[\frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI})+C_M\alpha_{AI}(1-\alpha_S)}, 1]$

Table A5 Second Opinion AI vs. Gatekeeper AI for $(1-\alpha_{AI})(1-\beta_{AI}) \leq (1-\alpha_S)(1-\beta_S),\ \beta_S > \beta_{AI}$**,** $\beta_{AI} < \alpha_S + \alpha_{AI}$ **and** $C_M \geq C_T$

$$D' = \left[(C_T\beta_{AI} + C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S) - C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})\right]^2 + 4C_M C_T \alpha_S(1-\alpha_S)(1-\alpha_{AI})^2\left[\beta_S(1-\beta_{AI})^2(1-\alpha_{AI}) - \beta_{AI}(1-\beta_S)^2(1-\alpha_S)\right]$$

$$C_S(p) - C_G(p) = \begin{cases} \frac{p(1-\alpha_S)[(1-p)C_T\beta_S(1-\beta_{AI})-pC_M\alpha_{AI}(1-\alpha_S)]}{(1-p)(1-\beta_{AI})}, & \text{if } p \leq p_1 \\ \frac{(1-p)(1-\beta_{AI})[(1-p)C_T\beta_S(1-\beta_{AI})-pC_M\alpha_{AI}(1-\alpha_S)]}{p(1-\alpha_S)}, & \text{if } p_1 < p < p_2 \\ \frac{(1-\alpha_S)(p((1-\alpha_{AI})+(1-\beta_S))-(1-\beta_S))(C_T(1-p)\beta_{AI}(1-\beta_S)-C_M p\alpha_S(1-\alpha_{AI}))}{p(1-\alpha_{AI})(1-\alpha_S)} & \\ +\frac{(1-p)(1-\alpha_{AI})(1-\beta_{AI})(C_T(1-p)\beta_S(1-\beta_{AI})-C_M p\alpha_{AI}(1-\alpha_S))}{p(1-\alpha_{AI})(1-\alpha_S)}, & \text{if } p \geq p_2. \end{cases}$$

Now, $C_S(p) - C_G(p) \leq 0$ if

$$p \geq \frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI}) + C_M\alpha_{AI}(1-\alpha_S)}, \quad p \leq \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)} \quad \text{or}$$
$$p \geq \frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI}) + C_M\alpha_{AI}(1-\alpha_S)}, \quad \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)} < p < \frac{(1-\beta_S)}{(1-\beta_S) + (1-\alpha_{AI})} \quad \text{or}$$
$$-C_M p^2\alpha_S(1-\alpha_S)(1-\alpha_{AI})^2 + C_T(1-p)^2\left[\beta_S(1-\beta_{AI})^2(1-\alpha_{AI}) - \beta_{AI}(1-\beta_S)^2(1-\alpha_S)\right]$$
$$+ p(1-p)\left[(C_T\beta_{AI} + C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S) - C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})\right] \leq 0,$$
$$p \geq \frac{(1-\beta_S)}{(1-\beta_S) + (1-\alpha_{AI})}.$$

Consider the quadratic function:

$$f(X) = -C_M\alpha_S(1-\alpha_S)(1-\alpha_{AI})^2X^2 + C_T\left[\beta_S(1-\beta_{AI})^2(1-\alpha_{AI}) - \beta_{AI}(1-\beta_S)^2(1-\alpha_S)\right] + X\left[(C_T\beta_{AI} + C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S) - C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})\right].$$

The quadratic function $f(X)$ is a downward-sloping parabola, which will certainly be negative as X (which equals $\frac{p}{1-p}$) approaches ∞ (equivalently, as p approaches 1). The discriminant of the quadratic $f(X)$ is

$$D' = \left[(C_T\beta_{AI} + C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S) - C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})\right]^2 + 4C_M C_T\alpha_S(1-\alpha_S)(1-\alpha_{AI})^2\left[\beta_S(1-\beta_{AI})^2(1-\alpha_{AI}) - \beta_{AI}(1-\beta_S)^2(1-\alpha_S)\right].$$

The vertex of the quadratic $f(X)$ is at X^* such that

$$X^* = \frac{(C_T\beta_{AI}+C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S)-C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})}{2C_M\alpha_S(1-\alpha_S)(1-\alpha_{AI})^2}.$$

The roots of $f(X)=0$, denoted by X_1^* and X_2^*, are

$$X_1^* = \frac{(C_T\beta_{AI}+C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S)-C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})-\sqrt{D'}}{2C_M\alpha_S(1-\alpha_S)(1-\alpha_{AI})^2}$$

$$X_2^* = \frac{(C_T\beta_{AI}+C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S)-C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})+\sqrt{D'}}{2C_M\alpha_S(1-\alpha_S)(1-\alpha_{AI})^2}.$$

- Case 1: If

$$\frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI})+C_M\alpha_{AI}(1-\alpha_S)} \geq \frac{(1-\beta_S)}{(1-\beta_S)+(1-\alpha_{AI})} \iff C_T\beta_S(1-\alpha_{AI})(1-\beta_{AI}) \geq C_M\alpha_{AI}(1-\alpha_S)(1-\beta_S),$$

then $C_S(p)-C_G(p)\geq 0$ for $p=p_2$. Further, $C_S(p)-C_G(p)\leq 0$ for $p=p_1$. Because $C_S(p)-C_G(p)$ is continuous, this implies $D'\geq 0$, and

$$C_S(p)-C_G(p)\leq 0 \iff$$
$$p\in\left[\frac{(C_T\beta_{AI}+C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S)-C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})+\sqrt{D'}}{(C_T\beta_{AI}+C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S)-C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})+\sqrt{D'}+2C_M\alpha_S(1-\alpha_S)(1-\alpha_{AI})^2},1\right].$$

An example where this case holds is $C_M=C_T=50$, $\alpha_{AI}=0.95$, $\beta_{AI}=0.81$, $\alpha_S=0.75$, $\beta_S=0.95$.

- Case 2: If

$$\frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI})+C_M\alpha_{AI}(1-\alpha_S)} < \frac{(1-\beta_S)}{(1-\beta_S)+(1-\alpha_{AI})} \iff C_T\beta_S(1-\alpha_{AI})(1-\beta_{AI}) < C_M\alpha_{AI}(1-\alpha_S)(1-\beta_S),$$

then $f\left(\frac{p}{1-p}\right)<0$ for $p=\frac{(1-\beta_S)}{(1-\beta_S)+(1-\alpha_{AI})}$. If the vertex of $f(X)$ is to the left of $\frac{(1-\beta_S)}{(1-\alpha_{AI})}$, then

$$\frac{\frac{(1-\beta_S)}{(1-\beta_S)+(1-\alpha_{AI})}}{1-\frac{(1-\beta_S)}{(1-\beta_S)+(1-\alpha_{AI})}} = \frac{(1-\beta_S)}{(1-\alpha_{AI})} \geq \frac{(C_T\beta_{AI}+C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S)-C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})}{2C_M\alpha_S(1-\alpha_S)(1-\alpha_{AI})^2},$$

which implies

$$C_M\alpha_{AI}(1-\beta_{AI}) \geq (C_T\beta_{AI}-C_M\alpha_S)(1-\beta_S),$$

then

$$C_S(p)-C_G(p)\leq 0 \iff p\geq \frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI})+C_M\alpha_{AI}(1-\alpha_S)}.$$

An example where this case holds is $C_M=C_T=50$, $\alpha_{AI}=0.90$, $\beta_{AI}=0.82$, $\alpha_S=0.60$, $\beta_S=0.95$.

Suppose now that the vertex is to the right:

$$\frac{\frac{(1-\beta_S)}{(1-\beta_S)+(1-\alpha_{AI})}}{1-\frac{(1-\beta_S)}{(1-\beta_S)+(1-\alpha_{AI})}} = \frac{(1-\beta_S)}{(1-\alpha_{AI})} < \frac{(C_T\beta_{AI}+C_M\alpha_S)(1-\alpha_{AI})(1-\alpha_S)(1-\beta_S)-C_M\alpha_{AI}(1-\alpha_S)(1-\alpha_{AI})(1-\beta_{AI})}{2C_M\alpha_S(1-\alpha_S)(1-\alpha_{AI})^2},$$

which implies

$$C_M\alpha_{AI}(1-\beta_{AI})+C_M\alpha_S(1-\beta_S) < C_T\beta_{AI}(1-\beta_S).$$

Because $1-\beta_{AI}>1-\beta_S$, if we restrict $C_M\geq C_T$, the above condition implies that $\beta_{AI}>\alpha_{AI}+\alpha_S$, which is not a practical assumption if all sensitivity and specificity values are greater than 0.5. *Q.E.D.*

Proof of Proposition 7: Follows directly from Propositions 5 and 6. Under the assumptions of Proposition 7, $\tau_G^a < \tau_S^G$ because

$$\alpha_{AI}\beta_{AI}(1-\alpha_S)(1-\beta_S) < \alpha_S\beta_S(1-\alpha_{AI})(1-\beta_{AI}) \iff$$
$$\frac{C_T\beta_{AI}(1-\beta_S)}{C_M(1-\alpha_{AI})\alpha_S + C_T\beta_{AI}(1-\beta_S)} < \frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI}) + C_M\alpha_{AI}(1-\alpha_S)}.$$

Further,

$$\frac{C_M}{C_T} > \frac{\beta_S}{\alpha_{AI}} \implies$$
$$\frac{C_T\beta_S(1-\beta_{AI})}{C_T\beta_S(1-\beta_{AI}) + C_M\alpha_{AI}(1-\alpha_S)} < \frac{(1-\beta_{AI})}{(1-\beta_{AI}) + (1-\alpha_S)} \leq \tau_G^c \text{ (see proof of Proposition 5)}.$$

Q.E.D.