Artificial Intelligence on Call: The Physician's Decision of Whether to Use AI in Clinical Practice*

Tinglong Dai Shubhranshu Singh

Abstract

Physicians are increasingly able to use artificial intelligence (AI) systems to aid their medical decision-making. This paper examines a physician's decision regarding whether to use an assistive AI system when prescribing a treatment plan for a patient. Using AI helps the physician generate an informative signal that lessens clinical uncertainty. It can also change the physician's legal liability in the event of patient harm. We analyze two patient-protection schemes that determine physician liability when using AI: the prevailing patient-protection scheme uses the AI signal to enforce the current standard of care, whereas an emerging scheme proposes using the AI signal as the new standard of care. We show that in both schemes, the physician has an incentive to use AI in low-uncertainty scenarios, even if AI provides little value. Furthermore, the physician may avoid using AI in higher-uncertainty scenarios where AI could have aided in better decision-making. As AI becomes more precise, the physician may become more hesitant to use it on certain patients. A comparison of the physician's decision to use AI under the two schemes reveals that using the AI signal as the new standard of care may mitigate AI underuse (overuse) for certain patients but may boost AI underuse (overuse) for some other patients.

Keywords: Artificial intelligence, treatment plan decisions, patient protection, clinical uncertainty

*Tinglong Dai is Professor of Operations Management and Business Analytics and Shubhranshu Singh is Associate Professor of Marketing, both at the Carey Business School, Johns Hopkins University. Both authors contributed equally to this work and are listed alphabetically. Email: dai@jhu.edu, shubhranshu.singh@jhu.edu. We thank Elodie Adida Goodman, W. Jason Choi, Jiatao Ding, Tony Ke, Qihong Liu, RJ Niewoehner, Dan Polsky, Jay Swaminathan, Risa Wolf, and Yue Wu for their helpful comments. We also received valuable feedback from seminar participants at Brandeis University; Chinese University of Hong Kong; Emory University; George Washington University; Harvard University; Institute for Mathematical and Statistical Innovation; Johns Hopkins University; The Ohio State University; Singapore Management University; Sungkyunkwan University; Temple University; Texas Christian University; University of Barcelona; University of California, Los Angeles; University of California, Riverside; University of Cambridge; University of Kansas; University of Miami; University of Nebraska–Lincoln; University of North Carolina at Chapel Hill; University of Pennsylvania; University of Southern California; and University of Toronto. Additionally, we benefited from interacting with session participants at 2022 INFORMS Marketing Science Conference, 2022 and 2023 Product and Service Innovation Conferences, 2022 INFORMS Annual Meeting, and 2022 POMS Annual Conference. This research was funded by the Johns Hopkins Discovery Award (2022-2024) and the Hopkins Business of Health Initiative's 2021 Pilot Grant Program.

In designing medical expert systems ... the actions should be thought of not as directly affecting the patient but as influencing the physician's behavior. If expert systems become reliably more accurate than human diagnosticians, doctors might become legally liable if they *don't* use the recommendation of an expert system.

Russell and Norvig (2015, p.1051), *Artificial Intelligence: A Modern Approach*

Introduction

Patients increasingly believe artificial intelligence (AI) can augment physicians' medical decisions and improve healthcare delivery (Khullar et al. 2022). To paraphrase Brynjolfsson and McAfee (2017), "AI won't replace physicians, but physicians who use AI will replace those who don't." AI has piqued the interest of the medical community since its inception in the 1950s. In recent years, departing from the initial focus on futuristic themes, AI has found concrete health applications of its time-tested techniques such as deep learning. For example, AI algorithms can assist clinicians in selecting treatment plans (i.e., medications and dosages) for oncology treatment (Price et al. 2019, Topol 2019b). In comparison to efforts to develop new AI tools, however, using existing AI tools in day-to-day healthcare practice remains uncommon (Topol 2019a), as reflected in a recent National Academy of Medicine report that laments the situation of "scant data describing successful clinical deployment of those models in health care settings" (Fihn et al. 2019). Mirroring this situation in practice, the literature has paid little attention to why and how physicians *use* medical AI in their medical practice—a perspective that we take in our research, which seeks to understand a physician's decision to use (or not to use) AI for different patients in response to the incentive and policy environments in which they are situated.[1]

In this paper, we develop a model of physicians' decision to use an assistive AI system in

[1] The medical literature on AI strategy has identified a number of challenges threatening its future, including AI algorithms' requirement of large-scale data and their lack of common sense (Rajkomar et al. 2019), their "black box" nature (Maddox et al. 2019), and the risk that they will amplify—rather than mitigate—human-induced biases (Obermeyer et al. 2019). However, such factors might be seen as product-development issues. Our focus is on physicians' decisions to *use* existing medical AI.

clinical practice in view of insurance reimbursement and potential liability.[2] As Harned et al. (2019) point out, "The introduction of novel medical technology into clinical practice gives rise to novel questions of legal liability when something goes wrong." Malpractice liability is a significant driver of the medical decision-making process even if AI is not an option (Baicker et al. 2007). When physicians have the option to use medical AI as part of a clinical encounter, they may be exposed to a higher or lower risk of malpractice liability in the event of patient harm; this risk, in turn, may cause them to use AI in ways that are not always conducive to patient health (Price et al. 2019), especially when a one-size-fits-all treatment plan does not suit all patients.

As a key motivating example, as described by Price et al. (2019), in the case of ovarian cancer treatment, a standard treatment plan entails administering 15 mg/kg of the chemotherapeutic bevacizumab every three weeks, whereas a nonstandard treatment plan can correspond to a higher dosage—75 mg/kg every three weeks. Although the standard treatment plan suits most patients, the nonstandard treatment plan may be optimal for those who require a greater dosage. Patients may be harmed if suboptimal treatment options are prescribed. Beyond clinical considerations, the physician's treatment-plan decision may be influenced by non-clinical considerations such as potential legal liability: when a physician deviates from the standard of care and an adverse patient outcome occurs, the practitioner is considered liable under the current patient-protection scheme (Price et al. 2019). In other words, when physicians prescribe the standard treatment plan, they are essentially immune from legal liability. As a result of such a patient-protection scheme, physicians may have an incentive to "play it safe" instead of prescribing the best treatment plan. Now that physicians have the option to use AI to supplement their judgment, how AI will influence their decision-making and—as often overlooked in the literature—if they will use AI in the first place is not immediately clear. This question prompts us to examine physicians'

[2] Liability determination can vary depending on whether the AI system is *autonomous* or *assistive* (Abràmoff et al. 2022a). In the case of *autonomous* AI, in which the AI system makes the eventual medical decision, the developer of AI is responsible for liability. Our focus in this paper is on *assistive* AI, whereby the physician makes the eventual medical decision and is thus responsible for liability.

decisions regarding the use of AI, including *whether* to use it and *how* to incorporate it into their practice. Note, these are two separate decisions: even when a physician opts to use AI, she may choose to follow or disregard its recommendations. The physician incurs no direct cost for disregarding an AI recommendation in the clinic environment, but she could face indirect costs due to potential liability implications, which are the focus of our paper.

We develop a theoretical model to examine a physician's decision of whether to use AI in the context of treatment-plan decision-making in view of different patient-protection schemes and insurance reimbursement rules. When the physician determines a personalized treatment plan for a patient, the treatment plan is subject to residual uncertainty in the sense that the physician is imperfectly informed about the optimal treatment plan for each patient. The physician can choose either the standard or nonstandard treatment plan that specifies the dosage of a given drug (e.g., chemotherapeutic bevacizumab for ovarian cancer treatment). Whereas the standard plan applies to many patients and helps shield the physician from legal liability when adverse patient outcomes occur, the nonstandard treatment plan may be optimal for certain patients. In addition to its clinical implications, prescribing the nonstandard treatment plan may generate higher revenue for the physician under a fee-for-service physician payment system. It may also create potential malpractice liability: the physician is shielded from legal liability when prescribing the standard treatment plan but may be liable when the nonstandard treatment plan is prescribed and patient harm occurs.

In the absence of the AI option, our analysis shows the amount of malpractice liability can be set to induce physicians to prescribe a treatment plan in the best interest of their patients. However, with the introduction of the option to use medical AI, how patient-protection schemes and reimbursement rules for AI use affect the physician's decision becomes less straightforward. Consistent with the notion that much of today's AI is essentially a "prediction machine" (Agrawal et al. 2018) and the thought framework proposed in a recent *JAMA* article by Price et al. (2019), we model the assistive AI system as an informational device that provides a predictive signal—which is informative but imperfect—indicating whether to prescribe the standard or nonstandard treatment plan.

The use of AI complicates patient-protection schemes, because whether the physician adheres to or deviates from the AI signal can have implications for patient outcomes. Two possible patient-protection schemes have been the topic of ongoing discourse about how incorporating AI into clinical practice will shape physician liability. These two schemes differ in terms of whether AI will override the standard of care, a distinction that has significant ramifications for physician liability in the event of patient damage.

We start by considering a patient-protection scheme, which, consistent with the prevailing legal practice (Tanenbaum et al. 2022), maintains that medical AI should be used to assist physicians in making better prescription decisions, not to replace the standard of care (Price et al. 2019). As Maliha et al. (2021) explain, "A physician who in good faith relies on an AI/ML system to provide recommendations may still face liability if the actions the physician takes fall below the standard of care... Physicians have a duty to independently apply the standard of care for their field, regardless of an AI/ML algorithm output." The AI system's recommendation can nevertheless be used as evidence in court to support the claims of either the physician or the patient in liability lawsuits. Accordingly, the prevailing practice implies that when the AI signal recommends the standard treatment plan but the physician deviates from it, the physician should face *more* liability for patient harm than when they do not use AI. Using the same rationale, when AI recommends the nonstandard treatment plan and the physician follows it, the physician assumes *less* liability.

One might anticipate that a physician will use AI when its signal can significantly reduce clinical uncertainty (i.e., when the physician is highly uncertain about the optimal treatment plan). Surprisingly, we find the physician may use AI in low-uncertainty cases (i.e., when the physician has little uncertainty about the optimal treatment plan) but avoid using it in higher-uncertainty cases. In low-uncertainty cases, the physician earns additional revenue from using AI but disregards its signal when determining the treatment plan. Due to the low uncertainty regarding the optimal treatment plan, the likelihood of AI returning a contradictory recommendation is low. Therefore, the expected liability for the physician is insignificant relative to the revenue gained from using AI. However, if the physician's uncertainty about the optimal treatment plan increases and the physician continues to dismiss the AI signal, the expected liability increases, discouraging the physician from using AI.

The preceding discussion demonstrates the physician has an incentive to overuse AI: when the physician is highly certain the standard (or nonstandard) treatment plan is right for the patient, she uses AI but then prescribes a treatment plan solely following her own assessment. Another way the physician overuses AI is by following its signal when her prior belief is precise enough to prescribe the treatment plan. This type of overuse may be observed when the physician is somewhat certain (but not highly certain) about the right treatment plan. In this case, adhering to the AI signal helps the physician mitigate potential liability.

Although the physician always overuses AI for some patients, we find non-clinical considerations (i.e., revenue and liability implications) may induce the physician to not use AI for certain patients who could have benefited from it. Specifically, the physician underuses AI when the additional revenue from prescribing the nonstandard plan is either very low or very high. If the additional revenue is very low, in light of liability consideration, the physician avoids prescribing the nonstandard plan. For this reason, the physician underuses AI for some patients by prescribing the standard treatment plan without using AI, thus eliminating the possibility of generating an AI signal that is indicative of the nonstandard treatment plan. On the other hand, if the additional revenue from the nonstandard treatment plan is very high, liability becomes less of the physician's concern. In this case, the physician underuses AI for some patients by prescribing the nonstandard treatment plan without using AI.

Policymakers and AI developers alike anticipate AI systems will be more widely used as they improve in accuracy (Edlich et al. 2019). We find, however, better AI accuracy can discourage the physician from using AI for certain patients. As AI becomes more accurate, the physician's expected liability increases if the physician uses AI but ignores its signal when making treatment decisions (e.g., when the physician is highly certain the nonstandard plan is the right plan for the patient). The physician becomes less likely to use AI in such situations. This result highlights an important point about AI use that is often ignored: policymakers should focus on incentivizing physicians to use AI appropriately but not necessarily more widely. Incentives aimed at broadening the use of AI may lead to overutilization.

The decision of a physician to use AI is also influenced by the insurance company's reimbursement policies. We consider two such policies in which the insurance company (1) reimburses the physician for using AI, or (2) only reimburses when the physician follows the treatment plan recommended by AI. In cases where the insurance company denies the claim, the patient is responsible for paying for AI. Our analysis shows the second reimbursement policy, in which the physician is only reimbursed for following AI's recommendations, leads to less overuse of AI and is preferred by the insurance company.

Next, we examine an emerging patient-protection scheme that echoes the view of the medical AI community (e.g., Grover 2019, Price et al. 2019, Russell and Norvig 2015, Sullivan and Schweikart 2019) that AI signals should replace the standard of care as they continue to improve in accuracy. This emerging scheme has become increasingly plausible "as AI/ML systems integrate into clinical care and become the standard of care" (Maliha et al. 2021). This emerging patient-protection scheme contends that the physician bears legal liability for patient harm if the physician deviates from what AI recommends. Specifically, the physician is liable for patient harm when (1) AI recommends the standard treatment plan but the physician chooses the nonstandard plan, or (2) AI recommends the nonstandard treatment plan but the physician chooses the standard plan. By contrast, the physician is *not* liable for patient harm when the physician prescribes what AI recommends, regardless of whether it is the standard or nonstandard treatment plan. Table 1 summarizes the key differences between the two liability schemes.

We find that, similar to the case of the first scheme, the physician has an incentive to use AI for low-uncertainty cases but may be reluctant to use it for higher-uncertainty cases. Both overuse and underuse of AI continue to exist. In addition, the result of the

Table 1: Comparison of the Two Patient-Protection Schemes

AI signal	**Physician decision**	**Is the physician liable for adverse outcomes?**	
		Prevailing scheme	**Emerging scheme**
Standard plan	Standard plan	No	No
Standard plan	Nonstandard plan	Yes	Yes
Nonstandard plan	Standard plan	No	Yes
Nonstandard plan	Nonstandard plan	Yes	No

physician being less likely to use a more precise AI system for some patients carries over. However, a comparison of the physician's decision to use AI under the two patient-protection schemes generates some important insights. First, moving from the first scheme to the second may reduce AI underuse for certain patients but boost AI underuse for some other patients. Second, the physician's tendency to overuse AI may decrease for some patients but increase for others. The implication is that no one scheme is universally superior to the other.

Our paper represents an initial modeling attempt to theorize how clinicians embed assistive AI systems into their clinical decision-making. Whereas enthusiasts believe AI has the potential to shape clinical practice in fundamental ways, we believe a more nuanced understanding of the endogeneity of physicians' use behavior is critical for medical AI's basic promises to be realized. Our findings demonstrate that how physicians use AI in their clinical practice is influenced by patient-protection schemes, in ways both non-obvious and sometimes detrimental. Accordingly, payers and policymakers need to focus on understanding physicians' corresponding use behavior in designing incentive and policy initiatives.

Literature

The expert-service literature on physicians' diagnostic and treatment decisions dates to Darby and Karni (1973), who propose that branding and client relationships may help reduce fraud committed by credence-goods sellers such as automobile mechanics and doctors. Dranove (1988) examines the physician's trade-off between revenue gains from unnecessary treatments and the negative effect these treatments have on reputation. Wolinsky (1993) shows consumer search and reputation considerations can help discipline experts in these markets. Durbin and

Iyer (2009) examine a corruptible doctor's incentive to advise the right treatment, and they show the introduction of side payments can facilitate truthful communication from the doctor to the patient. We contribute to this literature by modeling an expert's decision not only to acquire costly information, but also to *use* that information in diagnostic decision making. As a result, we identify scenarios in which the expert may knowingly acquire costly information even though he or she anticipate not incorporating it into the diagnostic decision.[3]

Our paper is related to a growing body of literature on incorporating AI into healthcare delivery. Longoni et al. (2019) find through a series of laboratory studies that patients may have a "uniqueness bias," believing AI is less likely to account for their unique qualities, posing a possible barrier to wider adoption of medical AI. Our paper complements Longoni et al. (2019) in deepening the line of inquiry related to barriers to adoption of AI, with the critical difference being that we focus on a barrier driven by healthcare practitioners' rational choice, rather than consumer psychology. Our approach reflects the unique characteristics of the healthcare industry, in which physicians, not patients, often decide whether to use medical AI (Abràmoff et al. 2022a, Russell and Norvig 2015). Dai and Singh (2020) develop a signaling model to show highly skilled physicians can underutilize diagnostic testing (or AI tool) to signal their skills. Our paper departs from Dai and Singh (2020) in two critical ways. On one hand, we focus on the case in which the physician uses AI in determining a treatment plan (for a patient with a diagnosed condition), not reaching a diagnostic decision. On the other hand, we focus on the impact of liability implications on the physician's decision to use AI, whereas Dai and Singh (2020) are primarily concerned with reputation signaling. In addition, our paper is thematically related to that of Kim et al. (2022), who examine the role of the *social planner* in shaping medical screening policies in the presence of physician liability considerations. However, our paper focuses on the *physician's* (as opposed to the

[3] Expert-service provision has been a topic of interest in the field of marketing. For example, Chandy et al. (2001) study the role of expert endorsers, Fitzsimons and Lehmann (2004) examine the adverse effect of unsolicited expert advice, and Schwartz et al. (2011) investigate the effect of the relationship with expert advisors. Shin (2007) considers how uninformed customers may consult an expert-service-providing retailer to resolve pre-purchase uncertainty. Singh (2017) examines a corruptible expert's incentives in the context of a procurement auction. Finally, Soberman (2009) looks at a media expert's strategies.

social planner's) discretionary use of AI in the context of two different patient protection schemes, which generate different implications for the physician's decision-not only whether to generate the AI signal, but also whether to follow the AI signal.

Our paper touches upon AI and health, two emerging themes in the marketing literature that have been addressed largely independently. In a non-health environment, Miklós-Thal and Tucker (2019) examine how better demand forecasting, as a result of improved machine-learning algorithms, influences how firms collude with each other in their pricing practices. Related to their counterintuitive finding that better algorithms do not necessarily make collusion behaviors more or less salient, we show better medical AI tools can lead to more salient AI avoidance behavior. Choi et al. (2021) study an emerging AI-enabled ship-then-shop model in which consumers, who subscribe to the service, make purchase decisions after the firm ships the product and consumers realize their fit. Liu et al. (2021) explore implications of social media platforms' use of AI algorithms in content moderation. Cao et al. (2021) examine the effect of competition on a firm's incentive to use an algorithm exploration and exploitation as opposed to one that only maximizes the quality of the next recommendation. Li et al. (2021) study a firm's decision to collect consumers' information to subsequently use it to set behavior-based pricing for consumers who have privacy concerns. Several theoretical studies (e.g., Amaldoss and He 2009, Bala et al. 2017, Dukes and Tyagi 2009, Jain and Li 2018, Jiang et al. 2014) have examined healthcare marketing decisions in terms of consumer preferences and welfare. Our paper is also related to the work by Zyung et al. (2020), who study a healthcare provider's decision to utilize an ethics committee when facing ethical dilemmas, Chen et al. (2020), who show the effectiveness of outreach marketing to encourage regular cancer screening, and Yu et al. (2022), who show money-back guarantees can serve as signals of unobservable expert-service quality in the context of in-vitro fertilization services. This paper combines the two themes and examines a physician's incentive to use (and not use) AI in clinical decision-making. We examine the physician's decision to use an AI system in the face of clinical uncertainty.

Model

Consider a patient who suffers from a medical condition that requires a course of treatment according to a physician-prescribed plan. A physician is in charge of prescribing and administering the treatment plan to the patient.[4] Consistent with the thought framework in Price et al. (2019), we consider the case in which two treatment plans, S and N, are available, where S represents a standard treatment plan and N represents a nonstandard treatment plan. A treatment plan is standard if it fits into the standard of care; conversely, a treatment plan is nonstandard if it deviates from the standard of care. Either the standard or the nonstandard plan can be the right plan for the patient. The probability that the standard plan is appropriate for the patient, unbeknownst to the physician until the physician-patient encounter, is $\alpha \in [0, 1]$; the nonstandard plan is the right plan with probability $(1 - \alpha)$. Determining α requires the physician's medical expertise and experience.[5] The physician must choose one of the two plans: $x \in \{S, N\}$. If the physician chooses the nonstandard treatment plan ($x = N$), relative to the standard treatment plan ($x = S$), she receives an additional revenue of r, reflecting higher administration fees that are often associated with higher dosages under prevailing physician reimbursement rules in the U.S., such as Medicare Part B coverage for prescription drugs administered in physician offices and hospital outpatient clinics (see, e.g., Gaynor et al. 2022).

After receiving the course of treatment, the patient's outcome, denoted by o, is either good (G) or bad (B). The outcome is good if the patient receives the right treatment, and bad otherwise. Thus, given α, if the physician prescribes the standard treatment plan ($x = S$), the result is good with probability α and bad with probability $(1 - \alpha)$. Similarly, a nonstandard plan ($x = N$) leads to the good outcome with probability $(1 - \alpha)$ and the bad outcome with probability α. The good outcome provides the patient a payoff $u > 0$, and the

[4] For brevity of exposition, in the rest of the paper, when we say a physician prescribes a treatment plan, we mean the physician prescribes and administers the treatment plan.

[5] In our model, patients do not have the expertise or ability to learn the value of α. Otherwise, the physician has no role in medical decision-making, because patients can determine their own treatment plans.

bad outcome leads to a harm $h > 0$.[6] Thus, given α, S gives the patient an expected utility of $\alpha u - (1-\alpha)h$, whereas N gives $(1-\alpha)u - \alpha h$.

If patient harm—in the amount of h—arises, the patient decides whether to file a liability lawsuit against the physician. The patient's disutility from filing a lawsuit is a draw from a uniform distribution $U[0, t]$ with cumulative distribution function (cdf) $F(\cdot)$. The outcome of the lawsuit depends on whether the physician followed the standard of care: the physician is protected against liability claims if the prescribed treatment plan is consistent with the standard of care. In this case, the patient does not receive any payout. However, in the case in which the prescribed treatment plan is nonstandard, the physician is held liable for patient harm, because "under current law, a physician faces liability only when she does not follow the standard of care and an injury results" (Price et al. 2019). If the court holds the physician liable for the patient's harm, the patient receives a payout of p_l and the physician incurs a cost of l_0. The patient's payout (p_l) and the physician's liability (l_0) may not be the same, because physicians are typically insured for malpractice liability suits.

The physician is *impurely* altruistic in the sense that she cares about both her (1) clinical (e.g., the patient's health outcome and cost) and (2) nonclinical objectives (e.g., own financial incentives and potential liability). Following the convention of the literature (e.g., Durbin and Iyer 2009, Gaynor, Mehta, and Richards-Shubik 2022), we assume the physician assigns a weight $\theta \geq 0$ to nonclinical objectives. When $\theta = 0$, the physician is fully altruistic such that she cares only about the patient's health outcome and costs when making clinical decisions. We assume the altruistic physician does not care about the patient's gain from the lawsuit payout. (The alternative assumption that an altruistic physician also cares about the patient's gain from the lawsuit payout can lead to unrealistic scenarios in which the altruistic physician intentionally causes harm to facilitate the payout to the patient.)

The physician has access to an AI system that can help her generate an informative but

[6] In practice, the harm can be different depending on whether (1) a patient requiring the standard treatment plan receives the nonstandard one or (2) a patient requiring the nonstandard treatment plan receives the standard one. For simplicity of analysis, we assume the harm is the same across both scenarios. Relaxing this assumption does not qualitatively change our main findings.

imperfect signal $\xi \in \{0, 1\}$, indicating which treatment plan is appropriate for the patient; a positive signal ($\xi = 1$) indicates the standard treatment plan, whereas a negative signal ($\xi = 0$) indicates the nonstandard treatment plan. The AI signal has a precision of $\rho \in \left(\frac{1}{2}, 1\right)$. As a feature unlike many conventional technologies, the black-box nature of an AI system, especially one built using complex neural networks, means that if something goes wrong, the system may "fail in ways that we simply don't understand" (Angwin 2023). If the physician receives a positive signal $\xi = 1$ from AI (i.e., AI recommends treatment plan S), she uses Bayes' rule to update the probability that treatment plan S will result in a good outcome ($o = G$), which can be written as

$$b_1(\alpha) = \frac{\alpha\rho}{\alpha\rho + (1-\alpha)(1-\rho)}.$$

At the same time, the physician forms an updated belief that treatment plan S will result in the bad outcome with the complementary probability. Similarly, when AI generates a negative signal $\xi = 0$ (i.e., AI recommends treatment plan N), the physician forms an updated belief that treatment plan S leads to the good outcome with probability

$$b_0(\alpha) = \frac{\alpha(1-\rho)}{\alpha(1-\rho) + (1-\alpha)\rho},$$

and the bad outcome with the complementary probability.

The patient incurs a cost of $c > 0$ when the physician uses AI. This cost is not necessarily monetary by nature. It can reflect the patient's potential privacy concerns (Price and Cohen 2019) and aversion to medical AI (Longoni et al. 2019). The cost c may also reflect the potentially growing cost of care due to increases in premiums, consistent with the longstanding "IT productivity paradox" in the healthcare sector (Jones et al. 2012). We assume $c < (2\rho - 1)(u + h)/2$; otherwise, the physician will not use AI for any patient.

Apart from its clinical implications, AI has an impact on the physician's non-clinical goals. The physician earns r_{AI} as a result of using AI. If the physician uses AI, she submits a claim of r_{AI} to the insurance company. The insurance company operates under one of the two possible reimbursement rules: (1) reimburse r_{AI} if the physician uses AI; and (2) reimburse r_{AI} if the physician uses AI to help determine a treatment plan (i.e., if the physician follows the AI signal).[7] The insurance company sets the reimbursement rule at the start of the game. In particular, the first reimbursement scheme we consider in our paper is aligned with the "per-use" reimbursement policy that insurance companies use to reimburse providers for using AI systems such as IDx-DR (an FDA-approved AI device used for diabetic retinopathy screening), and the provider is paid a fixed, predetermined fee each time the AI system is used (Abràmoff et al. 2022b). The second payment system we consider, on the other hand, is motivated by the New Technology Add-On Payment (NTAP) system, which provides an additional payment to a provider for the use of new technologies such as AI when there is a "significant clinical improvement." The NTAP system has been used to reimburse for the use of several FDA-approved AI systems (Parikh and Helmchen 2022). The objective of the insurance company is to set a rule that incentivizes the physician to use AI from the patient-surplus perspective. If the insurance company declines the physician's claim, the patient bears the cost r_{AI} (in addition to cost c) of the physician's AI use.

The physician is also concerned about the legal ramifications of not acting on the AI system's advice. On one hand, suppose the prevalent patient-protection scheme is the one in which AI is used to reinforce the standard of care. Under this scheme, if AI suggests the standard treatment plan (i.e., $\xi = 1$) but the physician deviates from the AI's signal (i.e., $x = N$), the patient faces a lower disutility of filing a lawsuit. The reason is that AI helps demonstrate the physician's accountability leading up to the nonstandard treatment plan. The level of decrease in the patient's disutility of filing a lawsuit (i.e., litigation cost) is denoted by $\triangle > 0$. The marginal patient, who is indifferent between whether or not to sue the physician for malpractice, has a disutility level of $p_l + \triangle$. Now consider that the

[7] If the insurance company never reimburses for the physician's AI use, the burden of paying r_{AI} falls on the patient. This case is essentially equivalent to the first reimbursement rule because r_{AI} can be absorbed in the constant c.

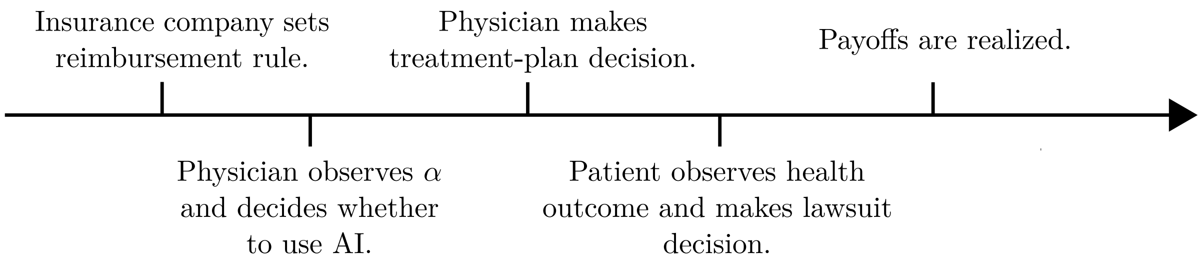


Figure 1: Timing of actions

patient suffers injury after AI suggests the nonstandard treatment plan and the physician follows it. We assume the patient's litigation costs will *increase* by $\triangle > 0$ in this situation, to account for the possibility that the physician will use AI to defend the chosen treatment plan in court. Therefore, the marginal patient has a disutility level of $p_l - \triangle$. On the other hand, under the emerging patient-protection scheme that uses AI signal as the standard of care, the physician is *not* liable if the physician follows the new standard of care, that is, AI's recommendation. However, if the physician does not follow AI's recommendation and patient harm occurs, the patient's disutility of filing a lawsuit is $p_l' > 0$, which can be less than p_l because AI may help substantiate the physician's responsibility. The AI signal (i.e., ξ) is assumed to be observable and verifiable in court.

The timing of actions, as illustrated shown in Figure 1, is as follows. First, the insurance company decides whether to reimburse r_{AI} whenever the physician uses AI or only when the physician decision is consistent with the AI recommendation ξ. Given the reimbursement rule and the patient's α, the physician decides whether to use AI to generate ξ. Next, the physician makes the treatment-plan decision x. Then, the patient's health outcome is realized and the patient decides whether to file a lawsuit. If the patient files a lawsuit, the court imposes a liability on the physician following the prevailing patient-protection scheme. In the last stage of the game, payoffs are realized. We solve the game given two different patient-protection schemes: (1) when AI is used to reinforce the standard of care, and (2) when AI signal overrides the standard of care. We also compare the physician's AI use decision corresponding to these patient-protection schemes.

Table 2: Summary of Key Notational Symbols

Symbol	Description
S	Treatment plan that fits the standard of care
N	Treatment plan that deviates from the standard of care
ξ	Signal from AI indicating appropriate treatment plan
ρ	Precision of AI signal
α	Prior likelihood that treatment plan S is the right one
r	Additional revenue physician earns if N is chosen
r_{AI}	Revenue physician earns from using AI
G	Good outcome for the patient
B	Bad outcome for the patient
u	Payoff for patient if outcome is good
h	Harm to patient if outcome is bad
c	Cost incurred by patient for AI use
t	Upper bound of uniform distribution $U[0,t]$
p_l	Payout for patient if physician is held liable
l_0	Cost for physician if held liable
l	Expected physician liability for patient harm if deviating from standard of care
l_{AI}	Expected change in physician liability for patient harm if deviating from standard of care (under the first patient-protection scheme)
l'	Expected physician liability for patient harm if deviating from AI recommendation (under the second patient-protection scheme)
θ	Weight physician assigns to nonclinical objectives

Before presenting the analysis, we summarize the main features of AI that we capture in our model. First, using AI directly increases the physician's payoff by r_{AI} and imposes a cost c on the patient. Second, following the AI signal can reduce treatment-plan selection errors, especially when the physician is highly uncertain about the correct plan. Third, using AI changes the probability that the physician will prescribe the non-standard plan, thus affecting the physician's revenue and expected liability. Fourth, the use of AI changes the physician's expected liability differently depending on the patient protection plan. We summarize our key notation in Table 2.

The Physician Decision in the Absence of AI

We start the analysis with the benchmark case in which the physician does not have access to the AI system. If the physician is fully altruistic and cares only about the patient's health outcome, she compares the patient's expected utility $\alpha u - (1-\alpha)\, h$ from plan S and $(1-\alpha)\, u - \alpha h$ from plan N. Comparing these expressions gives a threshold policy: the altruistic physician prescribes plan S if $\alpha \geq \frac{1}{2}$, and treatment plan N otherwise.[8]

If the physician is impurely altruistic, she also cares about her revenue r from the nonstandard treatment plan and expected liability from lawsuits. In the event of patient harm, the marginal patient who is indifferent between filing and not filing a lawsuit has an associated disutility $\hat{t} = p_l$, which translates into the physician's expected liability cost (given $x = N$ and patient harm) of

$$l \triangleq l_0 \cdot \int_0^{p_l} dF(t) = l_0 \cdot F(p_l).$$

The physician's expected payoff from prescribing treatment plans S and N are $\pi(\alpha|x = S) = \alpha u - (1-\alpha)h$ and $\pi(\alpha|x = N) = (1-\alpha)u - \alpha h + \theta(r - \alpha l)$. Comparing these payoffs gives rise to a threshold

$$\hat{\alpha}_0 \triangleq \frac{1 + \frac{\theta r}{u+h}}{2 + \frac{\theta l}{u+h}} \tag{1}$$

such that the physician prefers to prescribe the nonstandard treatment plan for patients with $\alpha < \hat{\alpha}_0$. A comparison of the physician's payoff from treatment plans S and N gives us the following lemma. (All proofs are in the Appendix.)

Lemma 1. *In the absence of AI, the physician prescribes treatment plan S if $\alpha \geq \hat{\alpha}_0$, and treatment plan N otherwise.*

The additional revenue r from treatment plan N induces the physician to prescribe it for some patients who should receive treatment plan S instead. Potential expected liability l mitigates the revenue-generation effect and induces the physician to move away from

[8] We assume the physician recommends treatment plan S when she is indifferent between recommending treatment plans S and N.

prescribing treatment plan N. The physician's expected payoff can be written as

$$\pi(\alpha) = \begin{cases} (1-\alpha)u - \alpha(h+\theta l) + \theta r & \text{if } \alpha < \hat{\alpha}_0, \\ \alpha u - (1-\alpha)h & \text{otherwise.} \end{cases} \quad (2)$$

We assume $u > \theta(r-l) - h$, which means that at $\alpha = 1$ (i.e., for a patient with a prior likelihood of requiring the standard treatment plan with certainty), the physician has no incentive to prescribe the nonstandard treatment. Stated differently, r is sufficiently low such that at least some patients will receive the standard treatment.

When AI Reinforces the Standard of Care

We now turn our attention to the use of AI in clinical practice. In this section, we consider the patient-protection scheme in which the AI signal is used to reinforce the standard of care. We start by analyzing the case in which the insurance company reimburses r_{AI} to the physician when she uses AI. We then analyze the physician's decision under the alternative reimbursement rule in which the insurance company reimburses r_{AI} to the physician when she not only uses AI but also follows the AI signal ξ to prescribe the treatment plan. Subsequently, we study the insurance company's reimbursement-rule decision.

Insurance Company Reimburses r_{AI} for AI Use

Consider that the insurance company reimburses r_{AI} when the physician uses AI. Because the focus of the healthcare community is on patient-welfare maximization, we start by analyzing a fully altruistic physician's AI-use decision. The following lemma specifies a fully altruistic physician's decision regarding whether to use AI:

Lemma 2. *A fully altruistic physician uses AI for patients with* α *satisfying* $(1-\rho) + \frac{c}{u+h} \leq \alpha < \rho - \frac{c}{u+h}$.

A fully altruistic physician uses AI only when the uncertainty about the optimal treatment plan is significant (i.e., when α is neither close to 0 nor 1). In such scenarios, the value of the AI-generated signal ξ justifies the patient's cost c. When opting to use AI, the altruistic physician *always* adheres to the AI signal. As the precision ρ of the AI signal increases, its use extends to a broader range of α values.

Next, we describe an *impurely* altruistic physician's expected payoff when augmented by AI. The physician observes the AI signal, $\xi \in \{0, 1\}$, and prescribes a treatment plan $x \in \{S, N\}$ for the patient. The physician faces liability when she prescribes plan N and patient harm occurs. Suppose the AI signal $\xi = 1$. If she deviates from the AI signal (i.e., prescribes plan N given $\xi = 1$) and patient harm arises, the patient who is indifferent between filing and not filing the lawsuit has a disutility level of $p_l + \triangle$. In this case, the physician's expected liability becomes

$$l_0 F(p_l + \triangle) = l_0 F(p_l) + l_0 [F(p_l + \triangle) - F(p_l)] = l + l_{AI},$$

where $l_{AI} \triangleq l_0 [F(p_l + \triangle) - F(p_l)]$. Now suppose the AI signal $\xi = 0$. If the physician follows the signal (i.e., prescribes plan N) and patient harm arises, the indifferent patient has a disutility of $p_l - \triangle$. In this case, the physician's expected liability is[9]

$$l_0 F(p_l - \triangle) = l_0 F(p_l) - l_0 [F(p_l) - F(p_l - \triangle)] = l - l_{AI}.$$

Because r_{AI} is paid by an insurance company rather than the patient, it does not directly affect the patient's utility. Nonetheless, the fee may have an *indirect* impact on patient utility by influencing the physician's decision-making process prior to prescribing the treatment plan. If AI advises the standard treatment plan and the physician follows through (i.e., $\xi = 1$ and $x = S$), the patient's expected utility is $b_1(\alpha)u - [1 - b_1(\alpha)]h - c$ and the

[9]Recall from the Model section that $F(\cdot)$ is the cdf of a uniform distribution. Thus, $F(p_l) - F(p_l - \triangle) = F(p_l + \triangle) - F(p_l)$. Our key insights carry over to the case in which the patient's disutility of filing a lawsuit does not follow a uniform distribution.

physician's expected payoff is $b_1(\alpha)u - [1 - b_1(\alpha)]h - c + \theta r_{AI}$. However, if AI advises the standard treatment plan and the physician chooses otherwise (i.e., $\xi = 1$ and $x = N$), the patient's expected utility is $[1 - b_1(\alpha)]u - b_1(\alpha)h - c$ and the physician's expected payoff is $[1 - b_1(\alpha)]u - b_1(\alpha)h - c + \theta\left[r + r_{AI} - b_1(\alpha)\left(l + l_{AI}\right)\right]$. A comparison of the physician's expected payoffs from $x = S$ and from $x = N$ (given $\xi = 1$) shows the physician prescribes $x = S$ for patients with $\alpha \geq \hat{\alpha}_1$, where

$$\hat{\alpha}_1 = \frac{(1-\rho)(u + h + \theta r)}{u + h - \theta(2\rho - 1)r + \theta\rho\left(l + l_{AI}\right)}.$$

Similarly, if AI advises the nonstandard treatment plan and the physician decides otherwise (i.e., $\xi = 0$ and $x = S$), the patient's expected utility is $b_0(\alpha)u - [1 - b_0(\alpha)]h - c$ and the physician's expected payoff is $b_0(\alpha)u - [1 - b_0(\alpha)]h - c + \theta r_{AI}$. However, if AI advises the nonstandard treatment plan and the physician acts upon it (i.e., $\xi = 0$ and $x = N$), the patient's expected utility is $(1 - b_0(\alpha))u - b_0(\alpha)h - c$ and the physician's expected payoff is $[1 - b_0(\alpha)]u - b_0(\alpha)h - c + \theta\left[r + r_{AI} - b_0(\alpha)\left(l - l_{AI}\right)\right]$. Therefore, in this scenario, the physician prescribes $x = S$ for patients with $\alpha \geq \hat{\alpha}_2$, where

$$\hat{\alpha}_2 = \frac{\rho(u + h + \theta r)}{u + h + \theta(2\rho - 1)r + \theta(1-\rho)\left(l - l_{AI}\right)}.$$

A comparison of thresholds $\hat{\alpha}_0$, $\hat{\alpha}_1$, and $\hat{\alpha}_2$ (presented in the following lemma) sheds light on the physician's treatment-plan decisions.

Lemma 3. $\hat{\alpha}_1 < \hat{\alpha}_0 < \hat{\alpha}_2$.

The physician disregards the AI signal and prescribes the nonstandard treatment plan ($x = N$) if the probability that plan S suits the patient is sufficiently low (i.e., $\alpha < \hat{\alpha}_1$). Similarly, the physician always prescribes the standard treatment plan ($x = S$) if α is sufficiently high (i.e., $\alpha \geq \hat{\alpha}_2$). When the uncertainty about whether the standard plan suits the patient is high, that is, in the intermediate range of α (i.e., $\hat{\alpha}_1 \leq \alpha < \hat{\alpha}_2$), the physician's prescription decision is consistent with the AI signal. As a result, an AI-augmented physician's expected payoff—which takes into account the probability of receiving signal $\xi = 1$, $\alpha\rho + (1-\alpha)(1-\rho)$, and the probability of receiving signal $\xi = 0$, $\alpha(1-\rho) + (1-\alpha)\rho$—can be summarized as

$$\pi_{AI}(\alpha) = \theta r_{AI} - c + \begin{cases} (1-\alpha)u + \theta r - \alpha\left[h + \theta\left(l + (2\rho-1)\, l_{AI}\right)\right] & \text{if } \alpha < \hat{\alpha}_1, \\ \rho u + \left[\alpha(1-\rho) + (1-\alpha)\rho\right]\theta r - (1-\rho)\left[h + \theta\alpha\left(l - l_{AI}\right)\right] & \text{if } \hat{\alpha}_1 \le \alpha < \hat{\alpha}_2\,, \\ \alpha u - (1-\alpha)h & \text{if } \alpha \ge \hat{\alpha}_2. \end{cases} \quad (3)$$

We now analyze an impurely altruistic physician's decision regarding whether to use AI. The physician's expected payoff from not using AI is described in (2), and that from using AI to assist in treatment plan decision-making is described in (3). The following proposition, derived from comparing (2) and (3), characterizes the physician's decision to use AI when the insurance company reimburses the physician on a per-use basis.

Proposition 1. *Under a reimbursement rule in which the physician receives r_{AI} for using AI, the physician's decision on whether to use AI is as follows:*

(i) If $\theta r_{AI} - c < -\hat{\alpha}_0\left[(2\rho-1)(u+h+\theta l-\theta r) + (1-\rho)\theta l_{AI}\right]$, the physician does not use AI for any patients.

(ii) If $-\hat{\alpha}_0\left[(2\rho-1)(u+h+\theta l-\theta r) + (1-\rho)\theta l_{AI}\right] \le \theta r_{AI} - c < 0$, the physician uses AI if and only if $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \le \alpha < \frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}$.

(iii) If $0 \le \theta r_{AI} - c < (2\rho-1)\theta l_{AI}\hat{\alpha}_1$, the physician uses AI if and only if $0 \le \alpha < \frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}$ or $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \le \alpha \le 1$.

(iv) If $\theta r_{AI} - c \ge (2\rho-1)\theta l_{AI}\hat{\alpha}_1$, the physician always uses AI for all patients.

An impurely altruistic physician, similar to a fully altruistic physician, does not use AI if the cost c to the patient is sufficiently high (i.e., when the condition in Proposition 1 *(i)* is satisfied). The physician makes this decision regardless of the degree of uncertainty (as reflected in α) about which treatment plan is ideal for the patient. Several factors influence her decision to use AI. A lower c induces her to use AI on more patients. Furthermore, for a given c, she is more likely to use AI if the revenue from using AI (r_{AI}) is higher or if the physician is less altruistic (i.e., θ is larger). Thus, the magnitude of $(\theta r_{AI} - c)$ dictates much of her AI-use decision; when this value is sufficiently low, she will not use AI for any patients.

Consider the case in which $\theta r_{\mathrm{AI}} - c < 0$ but is small, as per Proposition 1*(ii)*. In this situation, the physician does not use AI to aid the treatment-plan decision if she is fairly certain about the optimal treatment plan (i.e., if α is small or large enough). However, if she does not have enough certainty (i.e., for an intermediate range of α), she may use AI to generate an informative signal and then follows it in prescribing the treatment plan.

When $(\theta r_{AI} - c) > 0$, we make three observations about the physician's AI-use decision:

First, if $(\theta r_{AI} - c)$ is positive but not very large (i.e., $0 \leq \theta r_{AI} - c < (2\rho - 1)\theta l_{AI}\hat{\alpha}_1$), the physician will use AI even for patients for whom she is certain of the right treatment plan. Nonetheless, she chooses to not use AI for some of the patients whose treatment plan she is not certain of. The physician uses AI specifically for patients with $0 \leq \alpha < \frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}$ and $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha \leq 1$ but does not use it for patients with $\frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}} \leq \alpha < \frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}$. She prescribes plan N if $0 \leq \alpha < \frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}$. In this case, she receives an additional revenue of r_{AI} as a result of using AI. On the other hand, using AI increases her expected liability. When plan S is the right plan for the patient (i.e., with probability α), patient harm will occur (because the physician prescribes plan N) and the physician will face an additional liability of $\rho l_{AI} - (1-\rho)\, l_{AI} = (2\rho - 1)\, l_{AI}$. Note the physician is exposed to this additional risk with probability α. As a result, she is most inclined to use AI for patients with α close to zero. As α increases, whereas the added revenue from utilizing AI remains constant, the physician's expected liability increases, causing her to reevaluate whether to use AI. The physician eventually opts to not use AI for $\frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}} \leq \alpha < \frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}$. However, as α increases further, she is no longer certain about whether the treatment plan N suits the patient. Despite a different motivation, the physician contemplates using AI to mitigate clinical uncertainty. As α increases, she continues to use AI for all the patients with $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha \leq 1$. However, in this range of α, as α becomes higher than $\hat{\alpha}_2$, the physician's primary driver for using AI shifts from creating an informative signal to merely collecting the payment r_{AI}.

Second, the physician overuses AI, which occurs when she uses AI without intending to incorporate its signal. When the physician uses AI for patients with $\alpha < \hat{\alpha}_1$, she prescribes the nonstandard treatment plan regardless of the AI signal. Similarly, the physician uses AI for patients with $\alpha > \hat{\alpha}_2$ even when prescribing the standard treatment plan regardless of the AI signal. Non-clinical considerations drive the overuse of AI and reduce patient welfare.

Third, if $(\theta r_{AI} - c)$ is large enough (specifically, if $\theta r_{AI} - c \geq (2\rho - 1)\theta l_{AI}\hat{\alpha}_1$), the impurely altruistic physician uses AI for all the patients. This decision is in stark contrast to the fully altruistic physician's: the fully altruistic physician does not use AI for all the patients; rather, she uses AI only when doing so is in the best interest of patients. When $(\theta r_{AI} - c)$ is small but positive (i.e., $0 \leq \theta r_{AI} - c < (2\rho - 1)\theta l_{AI}\hat{\alpha}_1$), because of the tradeoff between the expected liability (which increases when a nonstandard plan is prescribed in conflict with $\xi = 1$) and revenue r_{AI} (which the physician receives for using AI), the impurely altruistic physician does not use AI for some patients. When $\theta r_{AI} - c \geq (2\rho - 1)\theta l_{AI}\hat{\alpha}_1$, the additional revenue r_{AI} becomes a more important consideration than the increased liability risk (caused by a relatively small l_{AI}), meaning she uses AI on all of the patients.

A policymaker may use r_{AI} as a means of incentivizing the physician to use AI. Therefore, understanding the effect of r_{AI} on the physician's decision to use AI is useful. Our analysis shows an increase in r_{AI} unambiguously increases the physician's incentive to use AI. The intuition is as follows: an increase in r_{AI} expands the range of patients for whom the physician will use AI and follow its recommendations (the upper bound of α in Proposition 1*(ii)* increases, while the lower bound decreases). Additionally, the physician is also more likely to use AI and ignore its recommendations. The threshold $\frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}$ in

Proposition 1*(iii)* increases with r_{AI}, whereas $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}$ decreases with r_{AI}.

Hereafter, we assume the physician uses AI for at least some patients; equivalently, the condition in Proposition 1*(i)* is not met. At this stage, the physician's non-clinical factors clearly sway the decision to use AI away from what is best for the patient. Two questions naturally arise: Do scenarios exist in which the physician *underuses* AI (i.e., the physician does not use AI when it is beneficial to the patient)? If so, how does her tendency to underuse AI change with her level of self-interest (as captured by θ)?

We now address these questions. The physician can underuse AI in two ways. On one hand, non-clinical concerns can lead her to prescribe the standard treatment plan without using AI (rather than use AI and follow its signal).[10] This condition can be expressed as

$$\frac{\partial}{\partial\theta}\left(\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}\right)<0,$$

which reduces to

$$r<\underline{r}\triangleq\frac{(1-\rho)\left(l-l_{AI}\right)\left[\rho\left(u+h\right)-c\right]-\left(u+h\right)r_{AI}}{2\rho\left(1-\rho\right)\left(u+h\right)+\left(2\rho-1\right)c}.$$

On the other hand, non-clinical concerns can lead the physician to prescribe the nonstandard treatment plan without using AI (rather than use AI and follow its signal). Equivalently,

$$\frac{\partial}{\partial\theta}\left(\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}\right)>0,$$

which reduces to

$$r>\bar{r}\triangleq\frac{\left[\rho l+\left(1-\rho\right)l_{AI}\right]\left[\left(1-\rho\right)\left(u+h\right)+c\right]+\left(u+h\right)r_{AI}}{2\rho\left(1-\rho\right)\left(u+h\right)+\left(2\rho-1\right)c}.$$

It is straightforward to show $\bar{r}>\underline{r}$.

The following proposition presents scenarios in which the physician underuses AI. In

[10] Recall that in Proposition 1*(ii)*, this tradeoff results in a lower bound of α.

the appendix, we define thresholds $\underline{\alpha}$ and $\bar{\alpha}$ to indicate the set of patients for whom the physician underuses AI when $r > \bar{r}$.

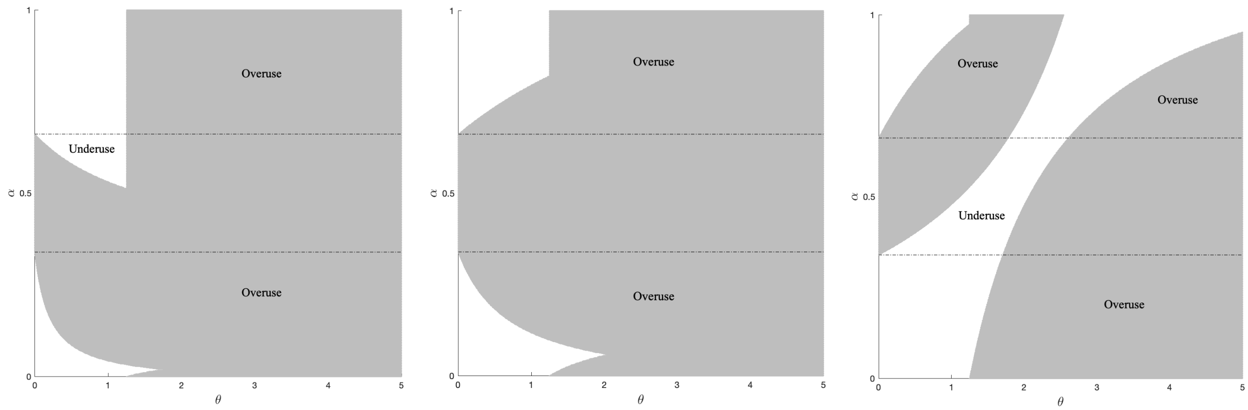
Proposition 2. *(i) If $r < \underline{r}$, the physician with $\theta < \frac{c}{r_{AI}}$ underuses AI for patients falling within the range of $\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})} < \alpha < \frac{\rho(u+h)-c}{u+h}$.*

(ii) If $\underline{r} \leq r \leq \bar{r}$, the physician does not underuse AI.

(iii) If $r > \bar{r}$, the physician underuses AI for $\underline{\alpha} < \alpha < \bar{\alpha}$.

Figure 2 illustrates these results. The proposition leads to an interesting observation that the physician may underuse AI when the additional revenue r from prescribing plan N is either small ($r < \underline{r}$) or big ($r > \bar{r}$), but not when it is in the intermediate range ($\underline{r} \leq r \leq \bar{r}$). Consider the case in which the physician is close to fully altruistic (i.e., θ is small). (We discuss the effect of varying θ in the next paragraph.) If r is small enough, the physician's expected revenue gain from prescribing plan N is minor relative to the liability risk. The physician's expected payoff increases as the physician prescribes plan S to more patients. The range of α for which the physician prescribes plan S expands. For some of the patients for whom the altruistic physician uses AI, the impurely altruistic physician essentially underuses AI by prescribing the standard treatment plan without using AI (and thus eliminating the possibility of generating a negative AI signal that advises the nonstandard treatment plan). According to a similar argument, when r is large enough, the impurely altruistic physician's revenue gains from prescribing plan N outweigh the liability implications. As a result, the physician's incentives to prescribe plan N increases. She underuses AI by prescribing the nonstandard treatment plan (without using AI) to some of the patients for whom using AI is welfare maximizing. The physician does not underuse AI in the intermediate range of r.

Another interesting observation from Proposition 2 is that the impact of physician self-interest (as measured by θ) on the extent of underusing AI (i.e., the range of α for which the physician underuses AI) can be non-monotonic. In particular, if θ is low, the extent of underusing AI (assuming it exists) increases with θ, because as θ increases, non-clinical



Case (i). $r < \underline{r}$ Case (ii). $\underline{r} \leq r \leq \bar{r}$ Case (iii). $r > \bar{r}$

FIGURE 2: Illustration of the three cases in Proposition 2. The physician uses AI in the shaded regions. Using AI is optimal from the patient's perspective between the two dotted lines.

considerations matter more to the physician. When θ is large enough, however, the extent of underusing AI weakly declines with θ because a large θ means the physician has a strong incentive to overuse AI. The physician uses AI out of revenue and liability considerations, albeit knowing ex ante that the AI signal will be ultimately dismissed. The overuse of AI spreads to a larger group of patients when θ is higher. For a large enough θ, the tendency for the physician to overuse AI prevails, and the extent of underusing AI eventually decreases, except when $r > \bar{r}$ and $\theta r_{AI} - c < [(1-\rho)(u+h)+c](2\rho-1)\theta l_{AI}/(u+h)$, in which case, the overuse of AI is not prevalent enough to affect AI underuse. The above discussion of physicians' underuse of AI implies the policy goal should not be to incentivize physicians to use AI, but rather to ensure they utilize it appropriately; a strategy aimed at promoting AI use may end up supplanting AI underuse with overuse.

Having examined the possibility of AI underuse, we turn to the physician's incentive to *overuse* AI. The above discussion indicates cases exist in which the physician overuses AI. However, whether the physician always overuses AI or only does so under specific circumstances (much like AI underuse) remains unclear. Whether the physician may utilize AI sparingly for certain patients while using it excessively for others is also unclear. The following proposition addresses these questions.

Proposition 3. *(i) If the physician deploys AI, she overuses it for patients with $\alpha \in A$,* *where* $A \subseteq \left[0, 1-\rho+\frac{c}{u+h}\right] \cup \left[\rho-\frac{c}{u+h}, 1\right]$.

(ii) When underusing AI for some patients, the physician overuses AI for some other patients.

If the physician's revenue gain r_{AI} is below the cost incurred to patient (c), and the difference between r_{AI} and c is sufficiently large, the physician does not use AI on any patients. However, if the physician uses AI at all, she overuses it for at least some patients. Recall the fully altruistic physician uses AI only for patients with α in the interval $[1-\rho+c/(u+h), \rho-c/(u+h)]$. An impurely altruistic physician uses AI for at least some patients outside of this interval. The overuse is motivated by the physician's desire to gain financially at the expense of patient welfare. The physician can overuse AI in two different ways: (1) by overusing AI and following its signal, and (2) by not only overusing AI but also ignoring its signal.[11] The first type of overuse is primarily driven by the physician's liability consideration and can be observed in physicians who are somewhat self-interested ($0 < \theta \leq c/r_{AI}$). By using AI and acting on its signal, the physician seeks to mitigate the expected liability in the event of patient harm. However, the second type of overuse, in which the physician ignores the AI signal, is primarily driven by her revenue consideration and requires her to be sufficiently motivated by self-interest ($\theta > c/r_{AI}$). Because overusing AI—with or without ignoring the AI signal—is detrimental to the patients, only those physicians who are insufficiently concerned about patient utility overuse AI. Non-clinical factors lead the physician to overuse AI for some patients, yet she does not always underuse AI. The implication is that when underusing AI for some patients, she overuses it for others.

Next, we examine the effect of AI precision on the physician's AI-use decision. AI developers and policymakers anticipate physicians will widely adopt AI to assist in clinical decision-making when AI precision becomes sufficiently high (Edlich et al. 2019). The following proposition offers a word of caution.

[11] Ignoring the AI signal does not imply the physician always chooses a different treatment plan than the AI signal suggests. It simply means that the physician chooses the same treatment plan (which may be either standard or nonstandard) regardless of the AI signal.

Proposition 4. *If* $0 \leq \theta r_{AI} - c < (2\rho - 1)\theta l_{AI}\hat{\alpha}_1$*, for patients with* $\alpha \in \left[0, \frac{\theta r_{AI} - c}{(2\rho-1)\theta l_{AI}}\right]$*, the physician is less likely to use AI as* ρ *increases.*

As ρ increases, that is, as AI becomes more precise, one may expect the physician to use it on more patients. (The altruistic physician, according to Lemma 2, does use AI on more patients as it becomes more precise.) Proposition 4 indicates the opposite may be true. Specifically, Proposition 4 means when the physician's gross margin from using AI (i.e., $\theta r_{AI} - c$) is positive but not sufficiently high, some α always exist in which the physician is less likely to use AI even as it becomes more precise. To understand why, note such α falls into the range of $[0, (\theta r_{AI} - c)/((2\rho - 1)\theta l_{AI})]$; the physician therefore always prescribes a nonstandard treatment for α within this range even when AI recommends a standard treatment. As AI becomes more precise, because α is close to zero, *ceteris paribus*, when patient harm arises, the physician is more likely to have deviated from AI recommendation. To avoid increased liability as a result of using AI, the physician avoids using AI.

Reimbursement Only When AI Helps in Decision-Making

Now, consider an alternative reimbursement rule whereby the insurance company only reimburses the physician for using AI when she follows the AI signal ξ to prescribe the treatment plan. In this case, if the insurance company declines the physician's claim, the burden of paying r_{AI} falls on the patient. If the physician follows the AI signal (i.e., if $x = S$ after receiving $\xi = 1$ or $x = N$ after receiving $\xi = 0$), both the patient's expected utility and the physician's expected payoff are the same as in the previous section. However, if the physician does not follow the AI signal ξ, the alternative reimbursement rule changes the tradeoffs. If the physician prescribes $x = N$ after receiving $\xi = 1$, the patient's expected utility is $[1 - b_1(\alpha)]\, u - b_1(\alpha)\, h - c - r_{AI}$ and the physician's expected payoff is $[1 - b_1(\alpha)]\, u - b_1(\alpha)\, h - c - r_{AI} + \theta\, [r + r_{AI} - b_1(\alpha)\,(l + l_{AI})]$. Similarly, if the physician prescribes $x = S$ after receiving $\xi = 0$, the patient's expected utility is $b_0(\alpha)\, u - [1 - b_0(\alpha)]\, h - c - r_{AI}$ and the physician's expected payoff is $b_0(\alpha)\, u - [1 - b_0(\alpha)]\, h - c - r_{AI} + \theta r_{AI}$.

Because the altruistic physician follows the AI signal when using AI, the range of α for which she uses AI is the same as in Lemma 2. Next, we examine the expected payoff of an impurely altruistic physician when augmented by AI. Suppose the AI signal suggests plan S (i.e., $\xi = 1$). A comparison of the physician's expected payoff from $x = S$ and $x = N$ shows she will prescribe plan S for patients with $\alpha \geq \check{\alpha}_1$, where

$$\check{\alpha}_1 = \frac{(1-\rho)\left(u+h+\theta r - r_{AI}\right)}{u+h-(2\rho-1)\left(\theta r - r_{AI}\right)+\theta\rho\left(l+l_{AI}\right)},$$

which is smaller than $\hat{\alpha}_1$. Similarly, after receiving $\xi = 0$, the physician prescribes plan S for patients with $\alpha \geq \check{\alpha}_2$, where

$$\check{\alpha}_2 = \frac{\rho\left(u+h+\theta r + r_{AI}\right)}{u+h+(2\rho-1)\left(\theta r + r_{AI}\right)+\theta\left(1-\rho\right)\left(l-l_{AI}\right)},$$

which is larger than $\hat{\alpha}_2$. Therefore, it is straightforward (from Lemma 3) that $\check{\alpha}_1 < \hat{\alpha}_0 < \check{\alpha}_2$.

A comparison of the physician's payoff from not using AI and from using AI reveals the physician's decision on whether to use AI. The following proposition describes the physician's AI use decision under the alternative reimbursement rule.

Proposition 5. *If the insurance company reimburses r_{AI} only when the physician follows the AI signal to prescribe the treatment plan, the physician's AI-use decision is the following:*

(i) If $\theta r_{AI} - c < -\hat{\alpha}_0\left[(2\rho-1)\left(u+h+\theta l-\theta r\right)+(1-\rho)\,\theta l_{AI}\right]$, the physician does not use AI for any patients.

(ii) If $-\hat{\alpha}_0\left[(2\rho-1)\left(u+h+\theta l-\theta r\right)+(1-\rho)\,\theta l_{AI}\right] \leq \theta r_{AI} - c < (1-\rho)\,r_{AI}$, the physician uses AI if and only if $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha < \frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}$.

(iii) If $(1-\rho)\,r_{AI} \leq \theta r_{AI} - c < \max\left\{(1-\rho)\,r_{AI}+(2\rho-1)\left(r_{AI}+\theta l_{AI}\right)\check{\alpha}_1, \rho r_{AI}-(2\rho-1)\,r_{AI}\check{\alpha}_2\right\}$, the physician uses AI if and only if $0 \leq \alpha < \frac{(\theta r_{AI}-c)-(1-\rho)r_{AI}}{(r_{AI}+\theta l_{AI})(2\rho-1)}$ or $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha < \frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}$ or $\frac{\rho r_{AI}-(\theta r_{AI}-c)}{(2\rho-1)r_{AI}} \leq \alpha \leq 1$.

(iv) If $\theta r_{AI} - c \geq \max\{(1-\rho) r_{AI} + (2\rho - 1)(r_{AI} + \theta l_{AI}) \check{\alpha}_1, \rho r_{AI} - (2\rho - 1) r_{AI} \check{\alpha}_2\}$*, the physician uses AI for all patients.*

Similar to Proposition 1, if $\theta r_{AI} - c$ is sufficiently small and negative, the physician does not use AI. When $\theta r_{AI} - c$ is relatively larger and satisfies the condition provided in Proposition 5 *(ii)*, the physician uses AI for patients when she is highly uncertain (i.e., in the intermediate range of α) and follows the AI signal to prescribe the treatment plan. When $\theta r_{AI} - c$ is larger than $(1-\rho) r_{AI}$, the following three observations highlighted in the discussion following Proposition 1 continue to hold: (1) The physician may use AI even for patients for whom she is certain of the right treatment plan, but not for those patients for whom she is more uncertain, (2) the physician overuses AI, and (3) when $\theta r_{AI} - c$ is sufficiently large, she uses AI for all patients. Because the intuition for these results is similar to those in Proposition 1, we do not repeat them here.

Next, we describe the insurance company's equilibrium reimbursement rule and highlight the key features of the physician's decision regarding whether to use AI.

Proposition 6. *In equilibrium, the insurance company reimburses the physician* r_{AI} *for using AI only when she follows the AI signal. As a result:*

(i) The physician may underuse AI if $r < \underline{r}$ *or if* $r > \bar{r}$ *but does not underuse AI if* $\underline{r} \leq r \leq \bar{r}$.

(ii) If the physician deploys AI, she overuses AI for at least some patients. When underusing AI for some patients, the physician overuses AI for other patients.

(iii) If $\frac{r_{AI}}{2} \leq \theta r_{AI} - c < \max\{(1-\rho) r_{AI} + (2\rho - 1)(r_{AI} + \theta l_{AI}) \check{\alpha}_1, \rho r_{AI} - (2\rho - 1) r_{AI} \check{\alpha}_2\}$*, the physician is less likely to use AI for* $0 \leq \alpha < \frac{(\theta r_{AI} - c) - (1-\rho) r_{AI}}{(r_{AI} + \theta l_{AI})(2\rho - 1)}$ *as* ρ *increases.*

If the physician intends to use AI and follow its signal to prescribe the treatment plan—specifically, when α is within the intermediate range specified in Proposition 5 *(ii)*—both reimbursement rules drive her AI-use incentives in the same way. The reason is that the physician receives the same payment r_{AI} directly from the insurance company under both reimbursement rules. If she disregards the AI signal (when α is sufficiently low or sufficiently high) and the insurance company denies her claim, the patient must pay r_{AI}. To the extent that the physician is concerned about the utility of the patient, the burden of paying r_{AI} reduces her incentive to use AI when it is not necessary. As a result, her inappropriate use of AI decreases. Because the physician is less likely to use AI when it is not desired by the patient, the insurance company reimburses her for AI use only when she follows the AI signal.

Although the equilibrium reimbursement rule reduces the likelihood of the physician ignoring the AI signal, it has no effect on her decision if she plans to follow the AI signal. As a result, similar to Proposition 2, the physician underuses AI when the revenue from the plan N (i.e., r) is sufficiently high (to avoid generating a $\xi = 1$ signal) and when it is sufficiently low (to avoid generating a $\xi = 0$ signal). When r is in an intermediate range, the physician does not underuse AI. Because the type of AI overuse in which the physician follows AI's signal exists when she uses AI for any patients, she overuses AI for some patients when she underuses it for some other patients. This result is similar to Proposition 3. Finally, generating an AI signal and ignoring it increases liability costs for the physician as AI becomes more precise. Therefore, similar to Proposition 4, as ρ increases, the physician become less likely to use AI for certain patients. The intuition for these results are provided in the previous section.

When AI Signal Is the Standard of Care

We now consider an emerging patient-protection scheme in which AI overrides the standard of care. This emerging scheme, which has been widely discussed in both the AI and medical communities (Russell and Norvig 2015, Sullivan and Schweikart 2019) contends that as it improves in precision, "AI becomes part of the standard of care, the consensus view of good medical practice," meaning "physicians may incur liability for rejecting correct but nonstandard AI recommendations and may conversely avoid liability for injury if they were following incorrect AI recommendations" (Price et al. 2019, p. 1766).

To be consistent with our modeling environment of the prevailing patient-protection scheme, we assume that when using AI, the physician is *not* liable if the physician follows the new standard of care, that is, AI's recommendation. the physician bears a liability when not following AI's recommendation and patient harm occurs. Because AI may help substantiate the physician's responsibility, the patient's expected disutility of litigation decreases. Therefore, the physician's expected liability, denoted by l' (and defined in the Online Appendix), is larger than l. All the other assumptions are the same as in the main model, where we considered the patient-protection scheme in which the AI reinforces the standard of care. Because the analysis and results are similar to those in the other patient-protection scheme, in this section, we provide an overview of the main results and highlight key differences. Formal statements of the physician's AI-use decision under both insurance reimbursement rules as well as their proofs are provided in the Online Appendix.

In the equilibrium, the insurance company's reimbursement rule requires the physician to follow the AI's signal. Otherwise, the insurance company does not reimburse r_{AI}. In line with Proposition 1, if the patient's cost c is sufficiently high, the physician does not use AI for any patients. Notably, the threshold c beyond which the physician refrains from using AI for any patients is higher under this emerging patient-protection scheme. This observation suggests the physician has a stronger incentive to use AI when its signal overrides the standard of care. In the following discussion, we consider the case in which the value of $(\theta r_{AI} - c)$ is large enough for the physician to use AI on at least some patients.

If $\theta r_{AI} - c < (1-\rho)\, r_{AI}$, the physician uses AI (and follows its signal) for patients for whom the physician is highly uncertain about the right treatment plan. In this situation, the emerging patient-protection scheme strictly expands the set of patients for whom the physician uses AI. An examination of the physician's decision to use AI in the case in which $(\theta r_{AI} - c)$ is larger than $(1-\rho)\, r_{AI}$ reveals the following three insights (mirroring those following Proposition 1): (1) If $(\theta r_{AI} - c)$ is positive but not too large, the physician uses AI for low-uncertainty cases (i.e., for patients with α close to 0 or 1) but avoids using AI for higher-uncertainty cases. (2) The physician overuses AI. (3) If $(\theta r_{AI} - c)$ is sufficiently large, the physician may use AI on all patients, regardless of their α. The intuition underlying these insights is similar to those described following Proposition 1. Despite the similarity across the physician's AI use decisions under the two patient-protection schemes, significant differences exist in the physician's decision to use AI, which we discuss in the following subsection.

Comparison of Patient-Protection Schemes

We now examine the differences between the two patient-protection schemes. To begin, we compare how the physician's tendency to overuse and underuse AI (i.e., deviation from patient-welfare maximizing use of AI) varies across these schemes. The thresholds $\underline{\alpha}'$ and $\bar{\alpha}'$ used in the following proposition are defined in the Appendix.

Proposition 7. *Under the emerging patient-protection scheme, the following properties hold:*

(i) The physician underuses AI for $\underline{\alpha}' < \alpha < \bar{\alpha}'$ *if* $r > \bar{r}' \triangleq \frac{l[c+(1-\rho)(u+h)]+(u+h)r_{AI}}{2\rho(u+h)(1-\rho)+(2\rho-1)c}$ *and does not underuse AI otherwise. The emerging patient-protection scheme leads to less underuse of AI if* $l' \leq \frac{(2\rho-1)l_{AI}+l}{\rho}$*, but more underuse for some patients otherwise.*

(ii) If the physician uses AI for any patients, the physician overuses it for at least some patients. The emerging patient-protection scheme leads to more AI overuse for $0 < \theta \leq 1 - \rho + \frac{c}{r_{AI}}$*. However, for* $\theta > 1 - \rho + \frac{c}{r_{AI}}$*, it leads to less overuse if* $l' > \frac{(2\rho-1)l_{AI}+l}{\rho}$*, and it leads to more overuse for some patients but less overuse for others otherwise.*

The physician underuses AI when prescribing a treatment plan (S or N) without using AI, due to non-clinical considerations. The emerging patient-protection scheme reduces the physician's incentive to prescribe plan S without using AI. As a result, the physician refrains from engaging in this form of underuse. This result is in sharp contrast to the finding presented in Proposition 2 where the physician underuses AI by prescribing plan S without using AI. However, similar to Proposition 2, if the physician's revenue r from plan N is sufficiently large, she underuses AI by prescribing plan N without using AI; her incentive to underuse AI depends on the magnitude of l'. As expected, when l' is small enough, non-clinical considerations increase the physician's willingness to use AI. As a result, she is less likely to underuse AI. On the other hand, when l' is large enough, the physician becomes less inclined to use AI; stated differently, non-clinical considerations induce her to increasingly underuse AI (by prescribing plan N without using AI) for certain patients.

If $(\theta r_{AI} - c)$ is negative and of a significant magnitude, the physician does not use AI with any patients. In Proposition 7*(ii)*, we consider the parameter space within which the physician uses AI for some patients. In this case, the physician overuses AI (regardless of the value of θ) for a subset of patients. The emerging patient-protection scheme incentivizes the physician to use AI and adhere to its signals when determining treatment plans. Although this practice may seem desirable, the concern is that the physician does so even for patients who may be properly treated without the use of AI. When non-clinical concerns are considerable, the physician also overuses AI by employing AI even when the treatment plan has been predetermined. Recall this type of AI overuse was also identified in Proposition 3. This observation prompts an immediate question: Is this emerging patient-protection scheme associated with an increase or decrease in such overuse of AI? The answer is situational. A significant liability l' reduces instances of AI overuse that occur when the physician uses AI but dismisses its signal. The rationale is that by following the AI signal rather than disregarding it, the physician might avoid a significant potential liability. However, if l' is small enough, the physician overuses AI more extensively to generate more revenue.

The emerging patient-protection scheme incentivizes more appropriate AI use and can potentially reduce AI underuse. These changes in the physician's AI use decision can lead to better clinical outcomes. However, the emerging scheme can also lead to more AI overuse for certain patients. These patients end up incurring the cost of using AI but receive no benefit. The emerging patient-protection scheme maybe more desired than the prevalent scheme if the patient's costs are relatively small and clinical benefits of using AI are significant. Next, we examine the possibility of the physician reducing the use of AI as ρ increases. Thresholds $\bar{m}$ and $\underline{m}$ are defined in the Online Appendix.

Proposition 8. *Under the emerging patient-protection scheme, if* $(1-\rho)\, r_{AI} \leq \theta r_{AI} - c < \max\{\bar{m}, \underline{m}\}$*, some* α *(which is either very low or very high) exists for which the physician is less likely to use AI as* ρ *increases.*

Mirroring Proposition 4, Proposition 8 states that under the emerging patient-protection scheme, scenarios exist in which the physician can be *less* likely to use AI as it improves in precision. This reduction in AI use corresponds to situations in which the physician uses AI but subsequently ignores its signal. Such behavior can be observed in both low- and high-α patients. By comparing Propositions 4 and 8, we show that although the emerging patient-protection scheme is intended to encourage the physician to use AI, it may unintentionally lead to increased AI avoidance even as AI improves in precision. The primary rationale behind this result is that the emerging patient-protection scheme relies on AI as the standard of care when the physician uses it; accordingly, using AI may lead to an increase in the physician's liability. Indeed, such AI avoidance may be better from the patient's perspective.

Extensions

In this section, we explore three extensions to our main model: first, the patient benefit of AI-assisted care, second, the data network effect, and third, the physician cost of improving AI accuracy. For each extension, we discuss the changes we make to our main model and how these considerations change our results. Proofs are provided in the online appendix.

Patient Utility from AI-Augmented Physician

In our main model, we assume the patient has chosen to see a physician who has access to an AI system for making treatment-plan decisions. Our main model views the patient as the "captive audience," as is the case in much of the prior literature. We will now consider an extension that addresses the patient's preference for the presence (vs. absence) of AI capabilities. This extension helps us understand the conditions under which the patient prefers the physician to have access to the AI system in making the treatment plan decision.

To simplify our analysis, we assume the patient belongs to either a high- or low-type. All else being the same, the standard treatment plan is more likely to be suitable for a high-type patient (H) than for a low-type patient (L). Prior to the patient's visit, the distribution of the patient's type (α) is specified as follows: $\Pr(\alpha = \alpha_H) = \beta$ and $\Pr(\alpha = \alpha_L) = 1 - \beta$. We assume $\beta > \frac{1}{2}$ to reflect the consideration that the majority of patients are of high type, such that the standard treatment plan applies to them. The patient's type is revealed to the physician only after a consultation and is known only to the physician. For simplicity of presentation, we assume the insurance company reimburses r_{AI} for the physician's AI use and the patient-protection scheme reinforces the standard of care.

We assume $\alpha_L < \hat{\alpha}_0 < \alpha_H$ so to focus on the interesting case in which without AI capabilities, the physician prescribes the standard treatment plan for high-type patients and the non-standard treatment plan for low-type patients. Thus, the patient's expected utility is

$$\begin{aligned}\mathbb{E}[\pi(\alpha)] &= \beta[\alpha_H u - (1-\alpha_H)h] + (1-\beta)[(1-\alpha_L)u - \alpha_L h] \\ &= [\alpha_H \beta + (1-\beta)(1-\alpha_L)]u - [\beta(1-\alpha_H) + (1-\beta)\alpha_L]h. \qquad (4)\end{aligned}$$

A comparison of the patient's utility from the physician with and without AI capabilities reveals that the patient prefers the physician with AI capabilities if

$$\rho > \max\{\alpha_H, 1-\alpha_L\} + \frac{c}{(1-\beta)(u+h)}.$$

The condition for the patient to prefer the physician to have AI capabilities is rather intuitive: all else equal, the patient prefers the physician to have access to AI if the accuracy of AI needs to exceed a threshold. This result qualitatively holds under the alternative reimbursement rule as well as under the emerging patient-protection scheme discussed in the main model. In addition, all the main insights continue to hold under the model setup of this extension.

Data Network Effect

AI accuracy is tied to the volume of data used for training the model. The more physicians adopt the use of AI in their practice, the more data the system can collect, leading to an increase in its accuracy. This phenomenon, known as the data network effect, states that the effectiveness of an AI system grows as it is able to learn from a larger pool of data. Varian (2019) argues the data network effect in the context of AI is essentially the well-known supply-side effect known as "learning by doing."

In line with the thought framework proposed by Varian (2019), we consider a model extension in which the physician "learns by doing" and derives utility from using AI. Following the literature on data network effects (Cabral 2011, Li et al. 2021, Lu et al. 2022), we capture the data network effect using a parameter η that represents the physician's "learning by doing" benefit from using AI. The physician's utility from using AI (when AI reinforces the standard of care and the insurance company always reimburses r_{AI} for the physician's AI use) can then be written as

$$\pi_{AI}(\alpha) = \theta r_{AI} + \eta - c + \begin{cases} (1-\alpha)u + \theta r - \alpha\left[h + \theta\left(l + (2\rho - 1)\, l_{AI}\right)\right] & \text{if } \alpha < \hat{\alpha}_1, \\ \rho u + \left[\alpha(1-\rho) + (1-\alpha)\rho\right]\theta r - (1-\rho)\left[h + \theta\alpha\left(l - l_{AI}\right)\right] & \text{if } \hat{\alpha}_1 \leq \alpha < \hat{\alpha}_2 \\ \alpha u - (1-\alpha)h & \text{if } \alpha \geq \hat{\alpha}_2. \end{cases} \tag{5}$$

By incorporating the data network effect through the parameter η, we confirm that all of our results are quantitatively valid. The data network effect, in essence, reduces the net cost of using AI and implies an intrinsic incentive for physicians to apply AI in the medical practice: it increases the likelihood of overuse of AI and decreases the likelihood of underuse. Stated differently, the data network effect serves as a strong motivator for physicians to utilize AI in their practice.

AI Precision at a Cost to the Physician

In this model extension, we consider the physician incurs a cost of using the AI system (e.g., to digitize, quantify, and standardize patients' information) which helps improve the AI system's precision ρ. We assume the physician incurs a cost of $k \cdot \rho$ when using the AI system, where $k > 0$ is a cost parameter. This cost specification captures the idea that a higher AI precision requires the physician to put more effort. We can write the physician's net revenue from using AI as $r_{AI} - k\rho$. A higher cost of increasing AI precision lowers the physician's payoff when using AI. The other assumptions are the same as in the main model.

We provide details of our analysis on pages A15–A17 of the online appendix. We find the physician's incentive to use AI is reduced due to the extra cost (relative to the main model) that the physician needs to incur to improve the AI precision. When the physician uses AI, she uses it for fewer patients. The overuse of AI decreases but the underuse expands. In addition, the result regarding the physician's reduced AI use increase in AI precision continues to exist. These results hold regardless of the insurance company's chosen reimbursement scheme. However, when the insurance company reimburses r_{AI} only when AI helps in decision making, the physician is less likely to reduce AI use with an increase in ρ. The insurance company prefers the reimbursement scheme in which the physician is expected to follow the AI signal in decision making. The results under the second patient-protection scheme, in which the AI signal is the standard of care, are qualitatively similar.

Concluding Remarks

One feature that distinguishes medical AI from other types of AI systems is that it is often experts (e.g., physicians), rather than end consumers (e.g., patients), who ultimately decide whether to use AI (Russell and Norvig 2015).[12] In this paper, we consider physicians who

[12] Our paper focuses on a type of medical AI that assists physicians in making medical decisions, such as screening, diagnosis, and treatment planning; other AI applications in healthcare settings, such as automatic transcription, patient monitoring, and process-flow automation, are beyond our scope.

choose treatment plans for patients, which entails the choice between the standard and nonstandard treatment plans. Physicians may use AI to generate an informative signal to supplement their evaluation. Our model captures the endogeneity of physicians' decision to use AI, which is driven by both clinical and non-clinical factors.

Our model focuses on the potential liability of using AI in healthcare, which is a distinctive feature of medical AI compared to conventional technologies. Whereas the liability of physicians using conventional technologies in their decision-making has been well explored, physician liability in cases where AI is used to augment healthcare is still the subject of debate (Price et al. 2019, Maliha et al. 2021). Another unique aspect of AI is that most diagnostic tests, such as blood tests, X-rays, and MRIs, are used to gather additional information rather than provide explicit medical recommendations. By contrast, AI systems often use existing patient information to make recommendations.

At a clinical level, physicians should conceivably use AI in high-uncertainty cases. However, potential malpractice liability is a significant factor in physicians' decision-making, and the use of AI perpetuates this concern. Our model accounts for physicians' potential liability based on their use of AI and how they make decisions in light of their own assessment and the AI signal. We demonstrate how complex trade-offs may lead physicians to use AI even under minimal uncertainty. Note that even when physicians use AI, they may not follow the AI signal. Thus, understanding both *when* and *how* physicians use AI is critical and leads to insights pertaining to when physicians may be overusing or underusing AI.

We show that under two proposed patient-protection approaches, physicians may engage in unnecessary use of AI in situations in which physicians' own information is sufficiently accurate to make treatment-plan decisions. In addition, physicians can avoid using AI in high-uncertainty scenarios out of liability concerns. As AI systems become more accurate, physicians may have a stronger tendency to avoid using AI for certain patients. Our findings underline the critical need to understand physicians' endogenous behavior when assessing the benefits of medical AI and associated patient-protection schemes.

Appendix

PROOF OF LEMMA 1. Suppose AI is not available to the physician. In this case, the physician's expected payoff from prescribing the standard plan $\pi(\alpha|x=S)$ is $\alpha u-(1-\alpha)h$. Similarly, the physician's expected payoff from prescribing the nonstandard plan $\pi(\alpha|x=N)$ is $(1-\alpha)u-\alpha h+\theta(r-\alpha l)$. Comparing these expected payoffs yields the lemma. *Q.E.D.*

PROOF OF LEMMA 2. A fully altruistic physician is only concerned about the patient's expected utility. In the absence of AI, the physician prescribes plan S if $\alpha\geq\frac{1}{2}$, and plan N otherwise. Therefore, in the absence of AI, the patient's expected utility is $\alpha u-(1-\alpha)h$ if $\alpha\geq\frac{1}{2}$, and $(1-\alpha)u-\alpha h$ otherwise. If the altruistic physician uses AI, the patient's expected utility depends on the recommendation from AI (i.e., $\xi\in\{0,1\}$) and the chosen treatment plan $x\in\{S,N\}$. Consider the following four cases:

(i) If $\xi=1$ and $x=S$, the patient's expected utility is $ub_1(\alpha)+(-h)[1-b_1(\alpha)]-c$.

(ii) If $\xi=1$ and $x=N$, the patient's expected utility is $(-h)b_1(\alpha)+u[1-b_1(\alpha)]-c$.

(iii) If $\xi=0$ and $x=S$, the patient's expected utility is $ub_0(\alpha)+(-h)[1-b_0(\alpha)]-c$.

(iv) If $\xi=0$ and $x=N$, the patient's expected utility is $(-h)b_0(\alpha)u+u[1-b_0(\alpha)]-c$.

A comparison of the patient's utilities in cases (i) and (ii) reveals the physician prescribes plan S upon receiving $\xi=1$ if $\alpha\geq1-\rho$. Similarly, the physician prescribes plan S upon receiving $\xi=0$ if $\alpha\geq\rho$. The physician prescribes plan N otherwise. Therefore, if AI is deployed, the patient's expected utility can be summarized as

$$U_{AI}=\begin{cases}(1-\alpha)u-\alpha h-c, & \text{if } \alpha<1-\rho,\\ \rho u-(1-\rho)h-c, & \text{if } 1-\rho\leq\alpha<\rho\,,\\ \alpha u-(1-\alpha)h-c, & \text{if } \alpha\geq\rho.\end{cases}$$

The physician uses AI if the patient's expected utility is higher in the presence of AI. A comparison of the patient's expected utility when AI is not deployed and when AI is deployed reveals the altruistic physician uses AI if $\frac{(1-\rho)(u+h)+c}{u+h} \leq \alpha < \frac{\rho(u+h)-c}{u+h}$. *Q.E.D.*

Proof of Lemma 3. We first show $\hat{\alpha}_1 < \hat{\alpha}_0$. Note that

$$\hat{\alpha}_0 - \hat{\alpha}_1 = \frac{(u+h+\theta r)}{2(u+h)+\theta l} - \frac{(1-\rho)(u+h+\theta r)}{u+h-(2\rho-1)\theta r+\rho\theta(l+l_{AI})} > 0$$

if $(2\rho-1)(u+h-\theta r+\theta l)+\rho\theta l_{AI} > 0$, which holds because $u+h-\theta r+\theta l > 0$.

Now we show $\hat{\alpha}_0 < \hat{\alpha}_2$. Note that $\hat{\alpha}_2 - \hat{\alpha}_0 = \frac{\rho(u+h+\theta r)}{u+h+(2\rho-1)\theta r+\theta(1-\rho)(l-l_{AI})} - \frac{u+h+\theta r}{2(u+h)+\theta l}$ is positive if $(2\rho-1)(u+h-\theta r+\theta l)+(1-\rho)\theta l_{AI} > 0$. Because $u+h-\theta r+\theta l > 0$, it is straightforward to show $\hat{\alpha}_0 < \hat{\alpha}_2$. As a result, $\hat{\alpha}_1 < \hat{\alpha}_0 < \hat{\alpha}_2$. *Q.E.D.*

Proof of Proposition 1. We start with establishing the following two claims.

Claim 1: Both $\pi(\alpha)$ and $\pi_{AI}(\alpha)$ are continuous in α.

In the absence of AI, the physician's payoff $\pi(\alpha)$, given in (2), is linearly decreasing in α for all $\alpha \in [0, \hat{\alpha}_0)$. The slope $\frac{\partial \pi(\alpha)}{\partial \alpha}\mid_{\alpha\in(0,\hat{\alpha}_0)} = -(u+h+\theta l)$. In the range $\alpha \in (\hat{\alpha}_0, 1]$, $\pi(\alpha)$ is linearly increasing and the slope is $\frac{\partial \pi(\alpha)}{\partial \alpha}\mid_{\alpha\in(\hat{\alpha}_0,1)} = u+h$. Because $\pi(\alpha)\mid_{\alpha\in(0,\hat{\alpha}_0)} = \pi(\alpha)\mid_{\alpha\in(\hat{\alpha}_0,1)}$ at $\alpha = \hat{\alpha}_0$, $\pi(\alpha)$ is continuous for any $\alpha \in [0,1]$. Similarly, $\pi_{AI}(\alpha)$, as given in (3), is continuous in the entire range of α because $\pi_{AI}(\alpha)\mid_{\alpha\in(0,\hat{\alpha}_1)} = \pi_{AI}(\alpha)\mid_{\alpha\in(\hat{\alpha}_1,\hat{\alpha}_2)}$ at $\alpha = \hat{\alpha}_1$, and $\pi_{AI}(\alpha)\mid_{\alpha\in(\hat{\alpha}_1,\hat{\alpha}_2)} = \pi_{AI}(\alpha)\mid_{\alpha\in(\hat{\alpha}_2,1)}$ at $\alpha = \hat{\alpha}_2$.

Claim 2: $\frac{\partial \pi_{AI}(\alpha)}{\partial \alpha}\mid_{\alpha\in(0,\hat{\alpha}_1)} < \frac{\partial \pi_{AI}(\alpha)}{\partial \alpha}\mid_{\alpha\in(\hat{\alpha}_1,\hat{\alpha}_2)} < 0 < \frac{\partial \pi_{AI}(\alpha)}{\partial \alpha}\mid_{\alpha\in(\hat{\alpha}_2,1)}$.

The first-order derivative of $\pi_{AI}(\alpha)$ can be written as

$$\frac{\partial \pi_{AI}(\alpha)}{\partial \alpha} = \begin{cases} -u-h-\theta\left[l+(2\rho-1)\,l_{AI}\right], & \text{if } \alpha \in (0, \hat{\alpha}_1), \\ -(2\rho-1)\,\theta r - \theta(1-\rho)(l-l_{AI}), & \text{if } \alpha \in (\hat{\alpha}_1, \hat{\alpha}_2), \\ u+h, & \text{if } \alpha \in (\hat{\alpha}_2, 1). \end{cases}$$

Given $u+h > \theta(r-l)$, Claim 2 follows.

Now we prove each statement in Proposition 1 one by one.

(i) If $\theta r_{AI} - c < 0$, $\pi(\alpha = 0) > \pi_{AI}(\alpha = 0)$. In addition, if $\alpha \in (0, \hat{\alpha}_1)$, $\frac{\partial \pi(\alpha)}{\partial \alpha} = -(u + h + \theta l) > \frac{\partial \pi_{AI}(\alpha)}{\partial \alpha} = -u - h - \theta [l + (2\rho - 1) l_{AI}]$; if $\alpha \in (\hat{\alpha}_2, 1)$, $\frac{\partial \pi(\alpha)}{\partial \alpha} = u + h = \frac{\partial \pi_{AI}(\alpha)}{\partial \alpha}$. Note $\pi(\alpha)$ is decreasing in α for $\alpha < \hat{\alpha}_0$ and increasing in α for $\alpha < \hat{\alpha}_0$. It is lowest at $\alpha = \hat{\alpha}_0$ and $\hat{\alpha}_1 < \hat{\alpha}_0 < \hat{\alpha}_2$. Therefore, to establish that the physician will not use AI, it suffices to show $\pi(\alpha = \hat{\alpha}_0) > \pi_{AI}(\alpha)|_{\alpha \in (\hat{\alpha}_1, \hat{\alpha}_2)}$, which is equivalent to $\theta r_{AI} - c < -\hat{\alpha}_0 [(2\rho - 1)(u + h + \theta l - \theta r) - (1 - \rho)\theta l_{AI}]$.

(ii) If $\theta r_{AI} - c < 0$, by comparing (2) and (3), we can see the physician does not benefit from using AI when $\alpha < \hat{\alpha}_1$ or $\alpha \geq \hat{\alpha}_2$. The physician can benefit from using AI only when $\hat{\alpha}_1 \leq \alpha < \hat{\alpha}_2$, in which case, $\pi_{AI}(\alpha) = \theta r_{AI} - c + \rho u + \theta [\rho - (2\rho - 1)\alpha] r - (1 - \rho) [h + \theta \alpha (l - l_{AI})]$. Comparing the above expression with (2) gives the condition for the physician to use AI:

$$\frac{(1 - \rho)(u + h + \theta r) - \theta r_{AI} + c}{u + h - \theta(2\rho - 1)r + \theta l_{AI} + \theta \rho (l - l_{AI})} \leq \alpha < \frac{\rho(u + h + \theta r) + \theta r_{AI} - c}{u + h + \theta(2\rho - 1)r + \theta(1 - \rho)(l - l_{AI})}. \tag{6}$$

In addition, $\hat{\alpha}_1 < \frac{(1-\rho)(u+h+\theta r) - \theta r_{AI} + c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} < \frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})} < \hat{\alpha}_2$.

(iii) and (iv) If $\theta r_{AI} - c \geq 0$, we derive the condition for the physician to use AI by examining the following three cases:

(a). If $\alpha < \hat{\alpha}_1$, by comparing $\pi(\alpha) = (1 - \alpha)u - \alpha(h + \theta l) + \theta r$ and $\pi_{AI}(\alpha) = \theta r_{AI} - c + (1 - \alpha)u + \theta r - \alpha [h + \theta (l + (2\rho - 1) l_{AI})]$, we find the condition for $\pi_{AI}(\alpha) \geq \pi(\alpha)$ is $0 \leq \alpha < \frac{\theta r_{AI} - c}{(2\rho - 1)\theta l_{AI}}$. However, to ensure $\frac{\theta r_{AI} - c}{(2\rho-1)\theta l_{AI}} < \hat{\alpha}_1$, we need $\theta r_{AI} - c < (2\rho - 1)\theta l_{AI} \hat{\alpha}_1$. If $\theta r_{AI} - c \geq (2\rho - 1)\theta l_{AI} \hat{\alpha}_1$, the physician uses AI for the entire range of $0 \leq \alpha < \hat{\alpha}_1$.

(b). If $\hat{\alpha}_1 \leq \alpha < \hat{\alpha}_2$, by comparing $\pi_{AI}(\alpha) = \theta r_{AI} - c + \rho u + \theta[\rho - (2\rho - 1)\alpha]r - (1 - \rho)[h + \theta \alpha (l - l_{AI})]$ with (2), we find the condition for $\pi_{AI}(\alpha) \geq \pi(\alpha)$ is the same as (6) above. However, because $\theta r_{AI} - c \geq 0$, we have $\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})} \geq \hat{\alpha}_2$. Therefore, if $\theta r_{AI} - c \geq 0$, the condition for the physician to use AI reduces to

$$\alpha \geq \frac{(1 - \rho)(u + h + \theta r) - \theta r_{AI} + c}{u + h - \theta(2\rho - 1)r + \theta l_{AI} + \theta \rho (l - l_{AI})}. \tag{7}$$

Furthermore, we can show

$$\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho\left(l-l_{AI}\right)}>\hat{\alpha}_1 \tag{8}$$

if and only if $\theta r_{AI}-c<(2\rho-1)\,\theta l_{AI}\hat{\alpha}_1$. In other words, if $\theta r_{AI}-c<(2\rho-1)\,\theta l_{AI}\hat{\alpha}_1$, the condition for the physician to use AI is (7); otherwise, the physician always uses AI for any α in the range of $\hat{\alpha}_1\le\alpha<\hat{\alpha}_2$.

(c). If $\alpha\ge\hat{\alpha}_2$, because $\pi_{AI}(\alpha)=\theta r_{AI}-c+\alpha u-(1-\alpha)h>\pi(\alpha)=\alpha u-(1-\alpha)h$, the physician always uses AI. *Q.E.D.*

Proof of Proposition 2. Recall that a physician who is interested in maximizing patient welfare uses AI for α, satisfying $\frac{(1-\rho)(u+h)+c}{u+h}\le\alpha<\frac{\rho(u+h)-c}{u+h}$.

(i) Consider the case presented in Proposition 1 (ii), where the physician with $\theta<\frac{c}{r_{AI}}$ uses AI for patients with $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}\le\alpha<\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}$.

If $r<\underline{r}$, we have $\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}<\frac{\rho(u+h)-c}{u+h}$ and $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}<\frac{(1-\rho)(u+h)+c}{u+h}$. Therefore, the physician underuses AI for patients with $\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}<\alpha<\frac{\rho(u+h)-c}{u+h}$. As θ increases beyond $\frac{c}{r_{AI}}$, the physician deploys AI for $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}\le\alpha\le1$ (see Proposition 1 (iii)). Therefore, if $r<\underline{r}$, no AI underuse occurs for $\theta\ge\frac{c}{r_{AI}}$.

(ii) If $\underline{r}\le r\le\bar{r}$, $\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}>\frac{\rho(u+h)-c}{u+h}$ and $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}<\frac{(1-\rho)(u+h)+c}{u+h}$. Therefore, no AI underuse exists in this region.

(iii) If $r>\bar{r}$, $\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}>\frac{\rho(u+h)-c}{u+h}$ and $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}>\frac{(1-\rho)(u+h)+c}{u+h}$. The physician underuses AI for patients with $\underline{\alpha}<\alpha<\bar{\alpha}$, where

$$\underline{\alpha}=\max\left\{\frac{(1-\rho)(u+h)+c}{u+h},\frac{\theta r_{AI}-c}{(2\rho-1)\,\theta l_{AI}}\right\},\text{ and}$$
$$\bar{\alpha}=\min\left\{\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})},\frac{\rho\,(u+h)-c}{u+h}\right\}.$$

If $\theta<\frac{c}{r_{AI}}$, $\frac{\partial}{\partial\theta}(\bar{\alpha}-\underline{\alpha})>0$; that is, the extent of AI underuse increases in θ when θ is small.

Note the physician overuses AI (i.e., the physician uses AI but ignores the AI signal) for

$\alpha < \frac{\theta r_{AI} - c}{(2\rho-1)\theta l_{AI}}$. If $\theta r_{AI} - c \leq \frac{(1-\rho)(u+h)+c}{u+h}(2\rho - 1)\,\theta l_{AI}$, the physician's AI underuse expands with θ and eventually covers the entire $\frac{(1-\rho)(u+h)+c}{u+h} \leq \alpha < \frac{\rho(u+h)-c}{u+h}$ region. Otherwise, in the region $\frac{(1-\rho)(u+h)+c}{u+h} < \alpha < \frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}$, the impurely altruistic physician overuses AI (whereas the altruistic physician engages in appropriate AI use). For a sufficiently large $(\theta r_{AI} - c)$ (i.e., $\theta r_{AI} - c > \frac{\rho(u+h)-c}{u+h}(2\rho - 1)\,\theta l_{AI}$), the physician's AI overuse expands to the entire $\frac{(1-\rho)(u+h)+c}{u+h} \leq \alpha < \frac{\rho(u+h)-c}{u+h}$ region, essentially eliminating all underuse. *Q.E.D.*

Proof of Proposition 3. An altruistic physician uses AI for patients with α satisfying $\frac{(1-\rho)(u+h)+c}{u+h} \leq \alpha < \frac{\rho(u+h)-c}{u+h}$. Here, we establish that an impurely altruistic physician ($\theta > 0$) overuses AI for some patients regardless of the value of θ.

Consider first that $\theta > \frac{c}{r_{AI}}$. Because a physician with $\theta > \frac{c}{r_{AI}}$ uses AI for $0 \leq \alpha < \frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}$, the physician overuses AI for patients with $0 \leq \alpha < \min\{\frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}, \frac{(1-\rho)(u+h)+c}{u+h}\}$.

Next, we establish that the physician with $0 < \theta \leq \frac{c}{r_{AI}}$ also overuses AI. If $r < \underline{r}$, $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} < \frac{(1-\rho)(u+h)+c}{u+h}$. In this case, the physician overuses AI for $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} < \alpha < \frac{(1-\rho)(u+h)+c}{u+h}$. If $\underline{r} \leq r \leq \bar{r}$, $\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})} > \frac{\rho(u+h)-c}{u+h}$ and $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} < \frac{(1-\rho)(u+h)+c}{u+h}$. In this case, the physician overuses AI for $\frac{\rho(u+h)-c}{u+h} < \alpha < \frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})} >$ and for patients with $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} < \alpha < \frac{(1-\rho)(u+h)+c}{u+h}$. Finally, if $r > \bar{r}$, $\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})} > \frac{\rho(u+h)-c}{u+h}$ and the physician overuses AI for patients with $\frac{\rho(u+h)-c}{u+h} < \alpha < \frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}$.

Clearly, the physician with $0 < \theta \leq \frac{c}{r_{AI}}$ also overuses AI for at least some patients. Therefore, an impurely altruistic physician always overuses AI for some patients.

Because a physician overuses AI for at least some patients regardless of the value of θ, it is straightforward that whenever the physician underuses AI for some patients, she simultaneously overuses AI for some other patients. *Q.E.D.*

Proof of Proposition 4. The condition $0 \leq \theta r_{AI} - c < (2\rho - 1)\theta l_{AI}\hat{\alpha}_1$ corresponds to case (iii) of Proposition 1. In this case, the physician uses AI if and only if $0 \leq \alpha < \frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}$ or $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha \leq 1$. As ρ increases, $\frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}$ decreases. Hence, the physician is less likely to use AI for some α that falls between 0 and $\frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}$.

Cases (i) and (iv) of Proposition 1 show the physician's AI-use decision does not change in ρ, because the former corresponds to "no use" and the latter corresponds to "always use."

For case (ii), that is, when $-\hat{\alpha}_0\left[(2\rho-1)\left(u+h+\theta l-\theta r\right)-(1-\rho)\,\theta l_{AI}\right]\leq\theta r_{AI}-c<0$, we know from Proposition 1 that the physician uses AI if and only if

$$\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}\leq\alpha<\frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}.$$

Observe that the right-hand side of the above inequality can be obtained by replacing ρ with $(1-\rho)$ on the left-hand side. Thus, if we take each side's first-order derivative in terms of ρ, these derivatives would have opposite signs. In other words, when the left-hand side decreases in ρ, the right-hand side must increase in ρ, and vice versa. For the physician's AI use to decrease in ρ for some θ, we need the right-hand side to decrease in ρ; that is, the right-hand side needs to have a negative first-order derivative in terms of ρ, which is equivalent to

$$(u+h)^2+\theta l(u+h)+\theta\{(c-r_{AI}\theta)[2r-(l-l_{AI})]+\theta r(r-l)\}<0. \tag{9}$$

If $l-l_{AI}\leq 2r$, we can show the left-hand side of (9) is always positive. Thus, for the above inequality to hold, a necessary condition is $l-l_{AI}>2r$; the condition means when the physician follows the AI's recommendation by choosing a nonstandard treatment plan, the physician's liability $(l-l_{AI})$—in case of patient harm—is more than twice the physician's additional revenue from the nonstandard treatment plan (r). Let us consider the case in which $l-l_{AI}>2r$. We can express the physician's margin from AI's recommendation as $\alpha(1-\rho)[r-(l-l_{AI})]+(1-\alpha)\rho\cdot r$, which can be rewritten as $\alpha[r-(l-l_{AI})]+\rho\{r+\alpha[(l-l_{AI}-2r)]\}$ and increases in ρ because $l-l_{AI}>2r$. In other words, as ρ increases, the physician's own utility increases. In addition, the patient's utility clearly increases in ρ. Thus, even under $l-l_{AI}>2r$, the physician is more likely to use AI as ρ increases. *Q.E.D.*

Proof of Proposition 5. The proof is similar to that of Proposition 1. A sketch of the proof is provided in the Online Appendix.

Proof of Proposition 6. Propositions 1 and 5 show the physician uses AI and follows ξ if $\theta r_{AI} - c < -\hat{\alpha}_0 \left[(2\rho - 1)(u + h + \theta l - \theta r) - (1 - \rho)\theta l_{AI}\right]$ for patients with

$$\frac{(1-\rho)(u+h+\theta r) - \theta r_{AI} + c}{u + h - \theta(2\rho - 1)r + \theta l_{AI} + \theta\rho\left(l - l_{AI}\right)} \leq \alpha < \frac{\rho(u+h+\theta r) + \theta r_{AI} - c}{u + h + \theta(2\rho-1)r + \theta(1-\rho)\left(l - l_{AI}\right)}$$

regardless of the reimbursement rule.

A comparison of the physician's inappropriate AI use (i.e., using AI even when she has pre-decided a treatment plan) decisions (part (iii) in Propositions 1 and 5) reveals she is more likely to overuse AI when the insurance company indiscriminately reimburses r_{AI}. Therefore, in equilibrium, the insurance company reimburses r_{AI} only when the physician follows ξ.

Because the range of α for which a physician uses AI and follows ξ to prescribe x is the same under the two reimbursement rules, part (i) follows from Proposition 2 and part (ii) follows from Proposition 3. The proof of part (iii) is similar to that of Proposition 4. *Q.E.D.*

References

Abràmoff MD, Cunningham B, Patel B, Eydelman MB, et al (2022a) Foundational considerations for artificial intelligence using ophthalmic images. *Ophthalmology* 129(2):e14–e32.

Abràmoff MD, Roehrenbeck C, Trujillo S, Goldstein J, et al (2022b) A reimbursement framework for artificial intelligence in healthcare. *npj Digital Medicine* 5(1).

Agrawal A, Gans J, Goldfarb A (2018) *Prediction Machines: The Simple Economics of Artificial Intelligence* (Cambridge, MA: Harvard Business Press).

Amaldoss W, He C (2009) Direct-to-consumer advertising of prescription drugs: A strategic analysis. *Marketing Science* 28(3):472–487.

Angwin J (2023) Autonomous vehicles are driving blind. *The New York Times* .

Baicker K, Fisher ES, Chandra A (2007) Malpractice liability costs and the practice of medicine in the medicare program. *Health Affairs* 26(3):841–852.

Bala R, Bhardwaj P, Chintagunta PK (2017) Pharmaceutical product recalls: Category effects and competitor response. *Marketing Science* 36(6):931–943.

Brynjolfsson E, McAfee A (2017) The business of artificial intelligence: What it can—and cannot—do for your organization. *Harvard Business Review* .

Cabral L (2011) Dynamic price competition with network effects. *Review of Economic Studies* 78(1):83–111.

Cao HH, Ma L, Ning ZE, Sun B (2021) How does competition affect exploration vs. exploitation? A tale of two recommendation algorithms. Working paper.

Chandy RK, Tellis GJ, Macinnis DJ, Thaivanich P (2001) What to say when: Advertising appeals in evolving markets. *Journal of Marketing Research* 38(4):399–414.

Chen Y, Lee JY, Sridhar SH, Mittal V, McCallister K, Singal AG (2020) Improving cancer outreach effectiveness through targeting and economic assessments: Insights from a randomized field experiment. *Journal of Marketing* 84(3):1–27.

Choi WJ, Liu Q, Shin J (2021) AI-driven retail transformation: Shop-then-ship or ship-then-shop? Working paper.

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

Darby MR, Karni E (1973) Free competition and the optimal amount of fraud. *Journal of Law Economics* 16(1):67–88.

Dranove D (1988) Demand inducement and the physician/patient relationship. *Economic Inquiry* 26(2):281–298.

Dukes AJ, Tyagi RK (2009) Pricing in vitro fertilization procedures. *Health Economics* 18(12):1461–1480.

Durbin E, Iyer G (2009) Corruptible advice. *American Economics Journal: Microeconomics* 1(2):220–42.

Edlich A, Phalin G, Jogani R, Kaniyar S (2019) Driving impact at scale from automation and AI. McKinsey & Company, URL https://mck.co/3IQIBVi.

Fihn S, Saria S, Mendonça E, Hain S, et al (2019) *Deploying AI in Clinical Settings*, chapter 6, 145–179 (Washington, DC: National Academy of Medicine).

Fitzsimons GJ, Lehmann DR (2004) Reactance to recommendations: When unsolicited advice yields contrary responses. *Marketing Science* 23(1):82–94.

Gaynor M, Mehta N, Richards-Shubik S (2022) Optimal contracting with altruistic agents: A structural model of medicare payments for dialysis drugs. NBER Working Paper 27172.

Grover A (2019) How one doctor does—and doesn't—use AI in his practice. *Wall Street Journal* .

Harned Z, Lungren MP, Rajpurkar P (2019) Machine vision, medical AI, and malpractice. *Harvard Journal of Law & Technology* .

Jain S, Li KJ (2018) Pricing and product design for vice goods: A strategic analysis. *Marketing Science* 37(4):592–610.

Jiang B, Ni J, Srinivasan K (2014) Signaling through pricing by service providers with social preferences. *Marketing Science* 33(5):641–654.

Jones SS, Heaton PS, Rudin RS, Schneider EC (2012) Unraveling the IT productivity paradox — lessons for health care. *New England Journal of Medicine* 366(24):2243–2245.

Khullar D, Casalino LP, Qian Y, Lu Y, et al (2022) Perspectives of patients about artificial intelligence in health care. *JAMA Network Open* 5(5):e2210309.

Kim Y, Ayvaci M, Raghunathan S, Ayer T (2022) When IT creates legal vulnerability: Not just overutilization but underprovisioning of health care could be a consequence. *MIS Quarterly* 45(3):1483–1516.

Li KJ, Li C, Zhang J (2021) Artificial intelligence: Information collection and behavior-based pricing under privacy concerns. Working paper.

Liu Y, Yildirim TP, Zhang ZJ (2021) Implications of revenue models and technology for content moderation strategies. Working paper.

Longoni C, Bonezzi A, Morewedge CK (2019) Resistance to medical artificial intelligence. *Journal of Consumer Research* 46(4):629–650.

Lu Z, Dou Y, Wu DJ, Chen J (2022) Selling smart and connected products: A value chain perspective. Working paper.

Maddox TM, Rumsfeld JS, Payne PRO (2019) Questions for artificial intelligence in health care. *JAMA* 321(1):31–32.

Maliha G, Gerke S, Cohen IG, Parikh RB (2021) Artificial intelligence and liability in medicine: Balancing safety and innovation. *Milbank Quarterly* 99(3):629–647.

Miklós-Thal J, Tucker C (2019) Collusion by algorithm: Does better demand prediction facilitate coordination between sellers? *Management Science* 65(4):1552–1561.

Obermeyer Z, Powers B, Vogeli C, Mullainathan S (2019) Dissecting racial bias in an algorithm used to manage the health of populations. *Science* 366(6464):447–453.

Parikh RB, Helmchen LA (2022) Paying for artificial intelligence in medicine. *npj Digital Medicine* 5(1):63.

Price WN, Cohen IG (2019) Privacy in the age of medical big data. *Nature Medicine* 25(1):37–43.

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

Rajkomar A, Dean J, Kohane I (2019) Machine learning in medicine. *New England Journal of Medicine* 380(14):1347–1358.

Russell SJ, Norvig P (2015) *Artificial Intelligence: A Modern Approach*. Prentice Hall Series in Artificial Intelligence (Upper Saddle River, NJ: Prentice Hall), 3rd edition.

Schwartz J, Luce MF, Ariely D (2011) Are consumers too trusting? the effects of relationships with expert advisers. *Journal of Marketing Research* 48(SPL):S163–S174.

Shin J (2007) How does free riding on customer service affect competition? *Marketing Science* 26(4):488–503.

Singh S (2017) Competition in corruptible markets. *Marketing Science* 36(3):361–381.

Soberman DA (2009) Marketing agencies, media experts and sales agents: Helping competitive firms improve the effectiveness of marketing. *International Journal of Research in Marketing* 26(1):21–33.

Sullivan HR, Schweikart SJ (2019) Are current tort liability doctrines adequate for addressing injury caused by AI? *AMA Journal of Ethics* 21(2):160–166.

Tanenbaum WA, Song K, Malek LA (2022) Theories of AI liability: It's still about the human element. *Reuters* .

Topol EJ (2019a) *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again* (Basic Books).

Topol EJ (2019b) High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine* 25(1):44–56.

Varian H (2019) *The Economics of Artificial Intelligence: An Agenda*, chapter 16: Artificial Intelligence, Economics, and Industrial Organization, 399–422 (Chicago, IL: University of Chicago Press).

Wolinsky A (1993) Competition in a market for informed experts' services. *RAND Journal of Economics* 24(3):380–398.

Yu S, Ghosh M, Viswanathan M (2022) Money-back guarantees and service quality: The marketing of in vitro fertilization services. *Journal of Marketing Research* 59(3):659–673.

Zyung JD, Mittal V, Kekre S, Hegde GG, Shang J, Marcus BS, Venkat A (2020) Service providers' decision to use ethics committees and consultation in complex services. *Journal of Marketing Research* 57(2):278–297.

Online Appendix to "Artificial Intelligence on Call: The Physician's Decision of Whether to Use AI in Clinical Practice"

Proof of Proposition 5. Because the proof is similar to that of Proposition 1, we only provide the main steps here. The physician's expected payoff is given by

$$\pi_{AI}(\alpha) = \theta r_{AI} - c + \begin{cases} (1-\alpha)u + \theta r - \alpha\left[h + \theta\left(l + (2\rho-1)\, l_{AI}\right)\right] - \left[\alpha\rho + (1-\alpha)\,(1-\rho)\right] r_{AI} & \text{if } \alpha < \check{\alpha}_1, \\ \rho u + \left[\alpha(1-\rho) + (1-\alpha)\rho\right]\theta r - (1-\rho)\left[h + \theta\alpha\left(l - l_{AI}\right)\right] & \text{if } \check{\alpha}_1 \le \alpha < \check{\alpha}_2, \\ \alpha u - (1-\alpha)h - \left[\alpha\,(1-\rho) + (1-\alpha)\,\rho\right] r_{AI} & \text{if } \alpha \ge \check{\alpha}_2. \end{cases} \tag{A1}$$

A comparison of equations (2) and (A1) characterizes the physician's decision to use AI. The continuity of the physician's payoff in α is established in the proof of Proposition 1. The physician's payoff $\pi_{AI}(\alpha)$ in equation (A1) in the entire range of α because

$$\pi_{AI}\left(\alpha\right)|_{\alpha\in(0,\check{\alpha}_1)} = \pi_{AI}\left(\alpha\right)|_{\alpha\in(\check{\alpha}_1,\check{\alpha}_2)} \text{ at } \alpha = \check{\alpha}_1, \text{ and}$$
$$\pi_{AI}\left(\alpha\right)|_{\alpha\in(\check{\alpha}_1,\check{\alpha}_2)} = \pi_{AI}\left(\alpha\right)|_{\alpha\in(\check{\alpha}_2,1)} \text{ at } \alpha = \check{\alpha}_2.$$

An examination of the first-order derivative of $\pi_{AI}(\alpha)$ reveals $\frac{\partial \pi_{AI}(\alpha)}{\partial\alpha}\,|_{\alpha\in(0,\check{\alpha}_1)} < \frac{\partial \pi_{AI}(\alpha)}{\partial\alpha}\,|_{\alpha\in(\check{\alpha}_1,\check{\alpha}_2)} < 0 < \frac{\partial \pi_{AI}(\alpha)}{\partial\alpha}\,|_{\alpha\in(\check{\alpha}_2,1)}$. To establish that the physician will not use AI, it suffices to show $\pi\left(\alpha = \hat{\alpha}_0\right) > \pi_{AI}\left(\alpha\right)|_{\alpha\in(\check{\alpha}_1,\check{\alpha}_2)}$, which is equivalent to $\theta r_{AI} - c < -\hat{\alpha}_0\left[(2\rho-1)\left(u + h + \theta l - \theta r\right) - (1-\rho)\,\theta l_{AI}\right]$, which proves part (i) of the proposition.

If $\theta r_{AI} - c < (1-\rho)\, r_{AI}$, the physician cannot benefit from using AI when $\alpha < \check{\alpha}_1$ or $\alpha \ge \check{\alpha}_2$. A comparison of $\pi_{AI}(\alpha)$ and $\pi\left(\alpha\right)$ in the $\check{\alpha}_1 \le \alpha < \check{\alpha}_2$ range reveals the physician uses AI for patients satisfying the condition

$$\frac{(1-\rho)(u+h+\theta r) - \theta r_{AI} + c}{u + h - \theta(2\rho-1)r + \theta l_{AI} + \theta\rho\left(l - l_{AI}\right)} \le \alpha < \frac{\rho(u+h+\theta r) + \theta r_{AI} - c}{u + h + \theta(2\rho-1)r + \theta(1-\rho)\left(l - l_{AI}\right)},$$

where $\check{\alpha}_1 < \frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}$ and $\check{\alpha}_2 > \frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}$.

If $\theta r_{AI} - c \geq (1-\rho) r_{AI}$, the condition for the physician's AI use depends on the following three cases:

(1) If $\alpha < \check{\alpha}_1$, the condition for $\pi_{AI}(\alpha) \geq \pi(\alpha)$ is $0 \leq \alpha < \frac{\theta r_{AI}-c-(1-\rho)r_{AI}}{(2\rho-1)(r_{AI}+\theta l_{AI})}$. However, to ensure $\frac{\theta r_{AI}-c-(1-\rho)r_{AI}}{(2\rho-1)(r_{AI}+\theta l_{AI})} < \check{\alpha}_1$, we need $\theta r_{AI} - c < (2\rho-1)(r_{AI}+\theta l_{AI})\check{\alpha}_1 + (1-\rho) r_{AI}$. If $\theta r_{AI} - c \geq (2\rho-1)(r_{AI}+\theta l_{AI})\check{\alpha}_1 + (1-\rho) r_{AI}$, the physician uses AI for the entire range of $0 \leq \alpha < \check{\alpha}_1$.

(2) If $\alpha \geq \check{\alpha}_2$, the condition for $\pi_{AI}(\alpha) \geq \pi(\alpha)$ is $\frac{\rho r_{AI}-(\theta r_{AI}-c)}{(2\rho-1)r_{AI}} \leq \alpha \leq 1$. However, to ensure $\frac{\rho r_{AI}-(\theta r_{AI}-c)}{(2\rho-1)r_{AI}} > \check{\alpha}_2$, we need $\theta r_{AI} - c < \rho r_{AI} - (2\rho-1) r_{AI}\check{\alpha}_2$. If $\theta r_{AI} - c \geq \rho r_{AI} - (2\rho-1) r_{AI}\check{\alpha}_2$, the physician uses AI for the entire $\check{\alpha}_2 \leq \alpha \leq 1$ region.

(3) If $\check{\alpha}_1 \leq \alpha < \check{\alpha}_2$, the condition for $\pi_{AI}(\alpha) \geq \pi(\alpha)$ is the same as part (ii) above.

Parts (iii) and (iv) follow from the above three cases. *Q.E.D.*

When AI Is the Standard of Care

Following a structure similar to the "When AI Reinforces the Standard of Care" section, consider first the reimbursement rule in which the insurance company reimburses for AI use whenever the physician uses AI.

In the event of patient harm, the marginal patient who is indifferent between filing and not filing a lawsuit has an associated disutility $\hat{t}' = p_l'$, which translates into the physician's expected liability cost (given $x = N$ and patient harm) of

$$l' \triangleq l_0 \cdot \int_0^{p_l'} dF(t) = l_0 \cdot F\left(p_l'\right).$$

We first consider the case in which AI recommends the standard treatment plan (i.e., $\xi = 1$). In this case, the physician's expected payoff from prescribing $x = S$ is $[b_1(\alpha) u - (1 - b_1(\alpha)) h - c] + \theta r_{AI}$, whereas the physician's expected payoff from prescribing

$x = N$ is $[-b_1(\alpha) h + (1 - b_1(\alpha)) u - c] + \theta (r_{AI} + r)$, where the terms inside the square brackets represent the patient utility and the terms outside the square brackets represent the physician's financial gain in each scenario. A comparison of the physician's expected payoff from $x = S$ and $x = N$ (given $\xi = 1$) reveals the physician prescribes $x = S$ for patients with $\alpha \geq \hat{\alpha}'_1$, where

$$\hat{\alpha}'_1 \triangleq \frac{(1-\rho)(u+h+\theta r)}{u+h+\theta[\rho l' - (2\rho-1)r]}.$$

Similarly, if the AI signal is $\xi = 0$, the physician's expected payoff from $x = S$ is $[b_0(\alpha) u - (1 - b_0(\alpha)) h - c] + \theta \{r_{AI} - [1 - b_0(\alpha)] l'\}$ and the physician's expected payoff from $x = N$ is $[-b_0(\alpha) h + (1 - b_0(\alpha)) u - c] + \theta (r_{AI} + r)$. As a result, given $\xi = 0$, the physician prescribes $x = S$ for patients with $\alpha \geq \hat{\alpha}'_2$, where

$$\hat{\alpha}'_2 \triangleq \frac{\rho[u+h+\theta(r+l')]}{u+h+\theta[(2\rho-1)r+\rho l']}.$$

The following lemma compares the magnitudes of $\hat{\alpha}'_1, \hat{\alpha}'_2$, and $\hat{\alpha}_0$.

Lemma A1. $\hat{\alpha}'_1 < \hat{\alpha}_0 < \hat{\alpha}'_2$

PROOF OF LEMMA A1. We first prove $\hat{\alpha}'_1 < \hat{\alpha}_0$ by noting

$$\begin{aligned}\hat{\alpha}_0 - \hat{\alpha}'_1 &= \frac{1 + \frac{\theta r}{u+h}}{2 + \frac{\theta l}{u+h}} - \frac{(1-\rho)(u+h+\theta r)}{u+h+\theta[\rho(l'-2r)+r]} \\ &= \frac{u+h+\theta r}{[2(u+h)+\theta l]\{u+h+\theta[\rho(l'-2r)+r]\}}[(u+h+\theta l-\theta r)(2\rho-1)+\theta\rho(l'-l)] > 0\end{aligned}$$

because $u + h + \theta l - \theta r > 0$ by assumption.

Next, we prove $\hat{\alpha}_0 < \hat{\alpha}'_2$. Because $\hat{\alpha}'_2 < 1$, we have

$$\hat{\alpha}'_2 = \frac{\rho[u+h+\theta(r+l')]}{u+h+\theta[\rho(2r+l')-r]} > \frac{\rho[u+h+\theta(r+l')] - \rho\theta l'}{u+h+\theta[\rho(2r+l')-r] - \rho\theta l'} = \frac{\rho(u+h+\theta r)}{u+h+\theta(2\rho-1)r}.$$

Then,

$$\frac{\rho(u+h+\theta r)}{u+h+\theta(2\rho-1)r} - \hat{\alpha}_0 = \frac{u+h+\theta r}{[2(u+h)+\theta l]\{u+h+\theta(2\rho-1)r\}} \cdot [(2\rho-1)(u+h-\theta r)+\theta\rho l],$$

which is positive because

$$(2\rho-1)(u+h-\theta r)+\theta\rho l > (2\rho-1)(-\theta l)+\theta\rho l = \theta l(1-\rho) > 0.$$

The proof is complete. *Q.E.D.*

The order of thresholds specified in the above lemma implies the physician ignores ξ and prescribes $x = N$ for all $\alpha < \hat{\alpha}'_1$, follows ξ for $\hat{\alpha}'_1 \leq \alpha < \hat{\alpha}'_2$, and ignores ξ and prescribes $x = S$ for all $\alpha \geq \hat{\alpha}'_2$. Therefore, under this emerging patient-protection scheme, the expected payoff of a physician who uses AI for all patients can be written as

$$\pi'_{AI}(\alpha) = \theta r_{AI} - c + \begin{cases} (1-\alpha)u - \alpha(h+\theta\rho l') + \theta r, & \text{if } \alpha < \hat{\alpha}'_1, \\ \rho u - (1-\rho)h + \theta[\rho - (2\rho-1)\alpha]r, & \text{if } \hat{\alpha}'_1 \leq \alpha < \hat{\alpha}'_2, \\ \alpha u - (1-\alpha)(h+\theta\rho l'), & \text{if } \alpha \geq \hat{\alpha}'_2. \end{cases} \tag{A2}$$

For ease of presentation, we define

$$\hat{\alpha}''_1 \triangleq \frac{c - \theta r_{AI} + (1-\rho)(u+h+\theta r)}{u+h+\theta l - \theta(2\rho-1)r} \text{ and } \hat{\alpha}''_2 \triangleq \frac{\theta r_{AI} - c + \rho(u+h+\theta r)}{u+h+\theta(2\rho-1)r}.$$

In addition,

$$\tilde{\alpha}_1 \triangleq \begin{cases} \frac{\theta r_{AI}-c}{\theta(\rho l'-l)} & \text{if } \rho l' \neq l \\ \hat{\alpha}'_1 & \text{if } \rho l' = l \end{cases} \text{ and } \tilde{\alpha}_2 \triangleq 1 - \frac{\theta r_{AI} - c}{\theta\rho l'}.$$

Note the definition of $\tilde{\alpha}_1$ hinges on whether $\rho l'$ is equal to l. If $\theta r_{AI} - c < -[(2\rho-1)(u+h-\theta r)+\theta\rho l]\hat{\alpha}_0$, the physician does not use AI. The following proposition, derived from comparing (A2) with (2), describes the physician's AI-use decision when using AI for at least some patients (i.e., if $\theta r_{AI} - c \geq \hat{\alpha}_0[u+h+(2\rho-1)\theta r] - \rho(u+h+\theta r)$).

Proposition A1. *Under the emerging patient-protection scheme, the physician's decision to use AI is characterized as follows:*

i. When $\rho l' \geq l$,

(a) if $-[(2\rho-1)(u+h-\theta r)+\theta\rho l]\hat{\alpha}_0 \leq \theta r_{AI} - c < 0$, the physician uses AI if and only if $\alpha \in [\hat{\alpha}_1'', \hat{\alpha}_2'']$;

(b) if $0 \leq \theta r_{AI} - c < \max\{\hat{\alpha}_1'\theta(\rho l' - l), (1-\hat{\alpha}_2')\theta\rho l'\}$, the physician uses AI if and only if $\alpha \in [0, \tilde{\alpha}_1] \cup [\hat{\alpha}_1'', \hat{\alpha}_2''] \cup [\tilde{\alpha}_2, 1]$;

(c) if $\theta r_{AI} - c \geq \max\{\hat{\alpha}_1'\theta(\rho l' - l), (1-\hat{\alpha}_2')\theta\rho l'\}$, the physician always uses AI.

ii. When $\rho l' < l$,

(a) if $-[(2\rho-1)(u+h-\theta r)+\theta\rho l]\hat{\alpha}_0 \leq \theta r_{AI} - c < \hat{\alpha}_1'\theta(\rho l' - l)$, the physician uses AI if and only if $\alpha \in [\hat{\alpha}_1'', \hat{\alpha}_2'']$;

(b) if $\hat{\alpha}_1'\theta(\rho l' - l) \leq \theta r_{AI} - c < 0$, the physician uses AI if and only if $\alpha \in [\tilde{\alpha}_1, \hat{\alpha}_2'']$;

(c) if $0 \leq \theta r_{AI} - c < (1-\hat{\alpha}_2')\theta\rho l'$, the physician uses AI if and only if $\alpha \in [0, \hat{\alpha}_2''] \cup [\tilde{\alpha}_2, 1]$;

(d) if $\theta r_{AI} - c \geq (1-\hat{\alpha}_2')\theta\rho l'$, the physician always uses AI.

Proof of Proposition A1. The proof entails comparing (A2) with (2) such that the physician uses AI when the former is greater than the latter. We consider the following four cases:

Case 1. $0 \leq \alpha < \hat{\alpha}_1'$. In this case, from (2) and (A2), we know the physician uses AI if and only if

$$\theta r_{AI} - c \geq \alpha\theta(\rho l' - l). \tag{A3}$$

i. If $\rho l' > l$,

(a) if $\theta r_{AI} - c < 0$, (A3) can never be satisfied;

(b) if $0 \leq \theta r_{AI} - c < \theta(\rho l' - l)\hat{\alpha}_1'$, (A3) is satisfied if $0 \leq \alpha < \frac{\theta r_{AI} - c}{\theta(\rho l' - l)}$;

(c) if $\theta r_{AI} - c \geq \hat{\alpha}_1' \theta(\rho l' - l)$, (A3) is always satisfied.

ii. If $\rho l' = l$,

(a) if $\theta r_{AI} - c < 0$, (A3) can never be satisfied;

(b) if $\theta r_{AI} - c \geq 0$, (A3) is always satisfied.

iii. If $\rho l' < l$,

(a) if $\theta r_{AI} - c < \theta(\rho l' - l)\hat{\alpha}_1'$, (A3) can never be satisfied;

(b) if $\hat{\alpha}_1' \theta(\rho l' - l) \leq \theta r_{AI} - c < 0$, (A3) is satisfied for $\frac{\theta r_{AI} - c}{\theta(\rho l' - l)} < \alpha < \hat{\alpha}_1'$;

(c) if $\theta r_{AI} - c \geq 0$, (A3) is always satisfied.

Case 2. $\hat{\alpha}_1' \leq \alpha < \hat{\alpha}_0$. In this case, from (2) and (A2), we know the physician uses AI if and only if

$$\theta r_{AI} - c \geq (1 - \alpha - \rho)(u + h) - \theta \alpha l + \theta[\alpha\rho + (1 - \alpha)(1 - \rho)]r. \tag{A4}$$

Note the right-hand side of (A4) decreases in α because its first-order derivative in terms of α is $-(u + h) - \theta l + (2\rho - 1)\theta r$, which is negative because it can be rewritten as

$$-(u + h) - \theta l + \theta r - 2(1 - \rho)\theta r < -(u + h) - \theta l + \theta r < 0 \text{ (using Assumption 1).}$$

We have three subcases:

i. If $\theta r_{AI} - c < (1 - \hat{\alpha}_0 - \rho)(u + h) - \theta\hat{\alpha}_0 l + \theta[\hat{\alpha}_0\rho + (1 - \hat{\alpha}_0)(1 - \rho)]r$, the physician never uses AI.

ii. If $(1-\hat{\alpha}_0-\rho)(u+h)-\theta\hat{\alpha}_0 l+\theta[\hat{\alpha}_0\rho+(1-\hat{\alpha}_0)(1-\rho)]r \leq \theta r_{AI}-c < (1-\hat{\alpha}_1'-\rho)(u+h)-\theta\hat{\alpha}_1' l+\theta[\hat{\alpha}_1'\rho+(1-\hat{\alpha}_1')(1-\rho)]r$, the physician uses AI if

$$\frac{(1-\rho)(u+h+\theta r)-(\theta r_{AI}-c)}{u+h+\theta[l-(2\rho-1)r]} \leq \alpha < \hat{\alpha}_0.$$

iii. If $\theta r_{AI}-c \geq (1-\hat{\alpha}_1'-\rho)(u+h)-\theta\hat{\alpha}_1' l+\theta[\hat{\alpha}_1'\rho+(1-\hat{\alpha}_1')(1-\rho)]r$, the physician always uses AI.

Case 3. $\hat{\alpha}_0 \leq \alpha < \hat{\alpha}_2'$. In this case, the physician uses AI if and only if

$$\theta r_{AI}-c \geq (\alpha-\rho)(u+h)-\theta[\alpha(1-\rho)+(1-\alpha)\rho]r. \tag{A5}$$

We can show the right-hand side of (A5) increases in α. Thus, we have three subcases:

i. If $\theta r_{AI}-c < (\hat{\alpha}_0-\rho)(u+h)-\theta[\hat{\alpha}_0(1-\rho)+(1-\hat{\alpha}_0)\rho]r$, the physician never uses AI.

ii. If $(\hat{\alpha}_0-\rho)(u+h)-\theta[\hat{\alpha}_0(1-\rho)+(1-\hat{\alpha}_0)\rho]r \leq \theta r_{AI}-c < (\hat{\alpha}_2'-\rho)(u+h)-\theta[\hat{\alpha}_2'(1-\rho)+(1-\hat{\alpha}_2')\rho]r$, the physician uses AI if

$$\hat{\alpha}_0 \leq \alpha < \frac{\rho(u+h+\theta r)+(\theta r_{AI}-c)}{u+h+\theta(2\rho-1)r}.$$

iii. If $\theta r_{AI}-c \geq (\hat{\alpha}_2'-\rho)(u+h)-\theta[\hat{\alpha}_2'(1-\rho)+(1-\hat{\alpha}_2')\rho]r$, the physician always uses AI.

Case 4. $\alpha \geq \hat{\alpha}_2'$. In this case, the physician uses AI if and only if

$$\theta r_{AI}-c \geq (1-\alpha)\theta\rho l'. \tag{A6}$$

We have the following three subcases:

i. If $\theta r_{AI}-c < 0$, (A6) can never be satisfied.

ii. If $0 < \theta r_{AI} - c < (1-\hat{\alpha}_2')\theta\rho l'$, (A6) is satisfied if $1 - \frac{\theta r_{AI}-c}{\theta\rho l'} \leq \alpha \leq 1$.

iii. If $\theta r_{AI} - c \geq (1-\hat{\alpha}_2')\theta\rho l'$, (A6) is always satisfied.

Combining cases 1–4 gives the proposition. Note that in combining cases 2 and 3, using (1), we obtain

$$
\begin{aligned}
&(\hat{\alpha}_0 - \rho)(u+h) - \theta[\hat{\alpha}_0(1-\rho) + (1-\hat{\alpha}_0)\rho]r \\
&= (1-\hat{\alpha}_0-\rho)(u+h) - \theta\hat{\alpha}_0 l + \theta[\hat{\alpha}_0\rho + (1-\hat{\alpha}_0)(1-\rho)]r \\
&= -[(2\rho-1)(u+h-\theta r) + \theta\rho l]\hat{\alpha}_0.
\end{aligned}
$$

We can thus use $-[(2\rho-1)(u+h-\theta r)+\theta\rho l]\hat{\alpha}_0$ to replace both $(\hat{\alpha}_0-\rho)(u+h)-\theta[\hat{\alpha}_0(1-\rho)+(1-\hat{\alpha}_0)\rho]r$ and $(1-\hat{\alpha}_0-\rho)(u+h)-\theta\hat{\alpha}_0 l+\theta[\hat{\alpha}_0\rho+(1-\hat{\alpha}_0)(1-\rho)]r$. *Q.E.D.*

Now, consider the reimbursement rule in which the insurance company reimburses for AI use only when the physician follows ξ. In this case, given the AI's signal $\xi = 1$, the physician prescribes plan S if $\alpha > \check{\alpha}_1' = \frac{(1-\rho)(u+h+\theta r-r_{AI})}{u+h+\theta\rho l'-(2\rho-1)(\theta r-r_{AI})}$, which is smaller than $\hat{\alpha}_1'$. In addition, given AI's signal of $\xi = 0$, the physician prescribes plan S if $\alpha > \check{\alpha}_2' = \frac{\rho\left(u+h+\theta r+r_{AI}+\theta l'\right)}{u+h+\theta\rho l'+(2\rho-1)(\theta r+r_{AI})}$, which is larger than $\hat{\alpha}_2'$.

The thresholds $\tilde{\alpha}_1'$ and $\tilde{\alpha}_2'$ are given by

$$\tilde{\alpha}_1' = \frac{\theta r_{AI} - c - (1-\rho)\, r_{AI}}{(\rho l' - l)\,\theta + (2\rho-1)\, r_{AI}}$$

and

$$\tilde{\alpha}_2' = \frac{\theta\rho l' + \rho r_{AI} - (\theta r_{AI} - c)}{\theta\rho l' + (2\rho-1)\, r_{AI}}.$$

For ease of presentation, we define $\bar{m} = \rho\,(\theta l' + r_{AI}) - \check{\alpha}_2'\,[\theta\rho l' + (2\rho-1)\,r_{AI}]$ such that $\theta r_{AI} - c < \bar{m}$ ensures $\tilde{\alpha}_2' > \check{\alpha}_2'$. Similarly, $\underline{m} = [(\rho l'-l)\,\theta + (2\rho-1)\,r_{AI}]\,\check{\alpha}_1' + (1-\rho)\,r_{AI}$ such that $\theta r_{AI} - c < \underline{m}$ ensures $\tilde{\alpha}_1' < \check{\alpha}_1'$. If $\theta r_{AI} - c < -\hat{\alpha}_0\,[(2\rho-1)\,(u+h-\theta r) + \rho\theta l]$, the physician does not use AI for any patients. The following proposition describes the physician's AI-use decision under this reimbursement rule when the physician uses AI at least for some patients.

Proposition A2. *If the insurance company reimburses r_{AI} only when the physician follows the AI signal to prescribe the treatment plan, the physician's AI use decision is the following:*

i. When $\rho l' \geq l$,

(a) if $-\hat{\alpha}_0 \left[(2\rho - 1)(u + h - \theta r) + \rho\theta l\right] \leq \theta r_{AI} - c < (1 - \rho) r_{AI}$, the physician uses AI if and only if $\alpha \in [\hat{\alpha}_1'', \hat{\alpha}_2'']$;

(b) if $(1 - \rho) r_{AI} \leq \theta r_{AI} - c < \max\{\bar{m}, \underline{m}\}$, the physician uses AI if and only if $\alpha \in [0, \tilde{\alpha}_1'] \cup [\hat{\alpha}_1'', \hat{\alpha}_2''] \cup [\tilde{\alpha}_2', 1]$;

(c) if $\theta r_{AI} - c \geq \max\{\bar{m}, \underline{m}\}$, the physician uses AI for all patients.

ii. When $\rho l' < l$,

(a) if $-\hat{\alpha}_0 \left[(2\rho - 1)(u + h - \theta r) + \rho\theta l\right] \leq \theta r_{AI} - c < \underline{m}$, the physician uses AI if and only if $\alpha \in [\hat{\alpha}_1'', \hat{\alpha}_2'']$;

(a) if $\underline{m} \leq \theta r_{AI} - c < (1 - \rho) r_{AI}$, the physician uses AI if and only if $\alpha \in [\tilde{\alpha}_1', \hat{\alpha}_2'']$;

(c) if $(1 - \rho) r_{AI} \leq \theta r_{AI} - c < \bar{m}$, the physician uses AI if and only if $\alpha \in [0, \hat{\alpha}_2''] \cup [\tilde{\alpha}_2', 1]$;

(d) if $\theta r_{AI} - c \geq \bar{m}$, the physician always uses AI.

Because the proof of Proposition A2 is similar to that of Proposition A1, we skip it. In addition, similar Proposition 6, we find that in the equilibrium, the insurance company sets the insurance rule in which the physician receives r_{AI} from the insurance company only when she follows ξ. It is straightforward to show the following:

(1) If $(\theta r_{AI} - c)$ is positive but not too large, the physician uses AI for low-uncertainty cases (i.e., for patients with α close to 0 or 1) but avoids using AI for higher-uncertainty cases.

(2) The physician overuses AI.

(3) If $(\theta r_{AI} - c)$ is sufficiently large, the physician may use AI on all patients, regardless of their α. *Q.E.D.*

PROOF OF PROPOSITION 7. (1) At $\theta = 0$, $\hat{\alpha}''_1 = \frac{c+(1-\rho)(u+h)}{u+h}$ and $\hat{\alpha}''_2 = \frac{-c+\rho(u+h)}{u+h}$. The physician's underuse of AI can be established by showing $\hat{\alpha}''_1 > \frac{c+(1-\rho)(u+h)}{u+h}$ or $\hat{\alpha}''_2 < \frac{-c+\rho(u+h)}{u+h}$ for $\theta > 0$. It is straightforward to show $\hat{\alpha}''_2 > \frac{-c+\rho(u+h)}{u+h}$ for all $\theta > 0$. However, $\hat{\alpha}''_1 > \frac{c+(1-\rho)(u+h)}{u+h}$ for $\theta > 0$, if $r > \bar{r}' \triangleq \frac{l[c+(1-\rho)(u+h)]+(u+h)r_{AI}}{2\rho(u+h)(1-\rho)+(2\rho-1)c}$. Therefore, the physician underuses AI for some patients, if $r > \bar{r}'$.

Next, we assume $r > \bar{r}'$ and identify the range of α ($\underline{\alpha}' < \alpha < \bar{\alpha}'$, where $\underline{\alpha}' \geq \frac{c+(1-\rho)(u+h)}{u+h}$ and $\bar{\alpha}' \leq \frac{-c+\rho(u+h)}{u+h}$) for which the physician underuses AI. In addition, we specify the set of physicians who underuse AI.

If $l \geq \rho l'$, the physician underuses AI for patients with $\underline{\alpha}' < \alpha < \bar{\alpha}'$, where $\underline{\alpha}' = \frac{c+(1-\rho)(u+h)}{u+h}$ and $\bar{\alpha}' = \min\left(\hat{\alpha}''_1, \tilde{\alpha}'_1\right)$. In this case, only physicians with $\theta \in (0, \theta')$ underuse AI, where $\theta' \leq 1 - \rho + \frac{c}{r_{AI}}$.

If $l < \rho l'$, $\underline{\alpha}' = \max\left(\frac{c+(1-\rho)(u+h)}{u+h}, \tilde{\alpha}'_1\right)$ and $\bar{\alpha}' = \min\left(\hat{\alpha}''_1, \frac{-c+\rho(u+h)}{u+h}\right)$. In this case, the range of θ (physicians who underuse AI) expands with l', and for a sufficiently large l' (which corresponds to $\tilde{\alpha}'_1 \leq \frac{c+(1-\rho)(u+h)}{u+h}$), all physicians with $\theta > 0$ underuse AI.

A comparison with expressions presented in the proof of proposition 2 reveals the following. Unlike in the case of the first patient-protection scheme, no AI underuse occurs when r is sufficiently small. If $l' \leq \frac{(2\rho-1)l_{AI}+l}{\rho}$, we have $\tilde{\alpha}'_1 \geq \frac{\theta r_{AI}-c-(1-\rho)r_{AI}}{(\rho l'-l)\theta+(2\rho-1)r_{AI}}$. In addition, $\hat{\alpha}''_1 < \check{\alpha}'_1$. Therefore, the extent of AI underuse is strictly lower if $l' \leq \frac{(2\rho-1)l_{AI}+l}{\rho}$. Now suppose $l' > \frac{(2\rho-1)l_{AI}+l}{\rho}$. In this case, $\tilde{\alpha}'_1 < \frac{\theta r_{AI}-c-(1-\rho)r_{AI}}{(\rho l'-l)\theta+(2\rho-1)r_{AI}}$, which implies the physician can potentially underuse AI for patients with $\tilde{\alpha}'_1 < \alpha < \frac{\theta r_{AI}-c-(1-\rho)r_{AI}}{(\rho l'-l)\theta+(2\rho-1)r_{AI}}$. Therefore, under the emerging patient-protection scheme, the physician may increasingly underuse AI if $l' > \frac{(2\rho-1)l_{AI}+l}{\rho}$.

(2) Note $\hat{\alpha}''_2 > \frac{-c+\rho(u+h)}{u+h}$ for all $\theta > 0$. In addition, $\tilde{\alpha}'_1 > 0$, for all $\theta > 1 - \rho + \frac{c}{r_{AI}}$. Therefore, overuse of AI exists when the physician uses AI.

If $0 < \theta < 1 - \rho + \frac{c}{r_{AI}}$, under the first patient-protection scheme, the physician uses AI for patients with $\alpha \in [\hat{\alpha}'_1, \hat{\alpha}'_2]$. Because $\hat{\alpha}''_1 < \hat{\alpha}'_1$ and $\hat{\alpha}''_2 > \hat{\alpha}'_2$ for all $\theta > 0$. It follows that the emerging patient-protection scheme strictly increases the overuse of AI, for $0 < \theta < 1 - \rho + \frac{c}{r_{AI}}$ regardless of the value of l'.

Now suppose $\theta > 1 - \rho + \frac{c}{r_{AI}}$. In this case, under the emerging patient-protection scheme, the physician does not use AI for patients with $\alpha \in [\hat{\alpha}''_2, \tilde{\alpha}'_2]$. However, the physician uses AI for these patients under the first patient-protection scheme. Therefore, the extent of overuse decreases for these patients. In addition, if $l' > \frac{(2\rho-1)l_{AI}+l}{\rho}$, $\tilde{\alpha}'_1 < \frac{\theta r_{AI}-c-(1-\rho)r_{AI}}{(\rho l'-l)\theta+(2\rho-1)r_{AI}}$. In other words, the physician does not use AI for patients with $\alpha \in \left[\tilde{\alpha}'_1, \frac{\theta r_{AI}-c-(1-\rho)r_{AI}}{(\rho l'-l)\theta+(2\rho-1)r_{AI}}\right]$. The implication is that AI overuse is less likely to occur.

Finally, we show the overuse of AI can increase for some patients if $\theta > 1 - \rho + \frac{c}{r_{AI}}$ and $l' < \frac{(2\rho-1)l_{AI}+l}{\rho}$. In this case, $\tilde{\alpha}'_1 > \frac{\theta r_{AI}-c-(1-\rho)r_{AI}}{(\rho l'-l)\theta+(2\rho-1)r_{AI}}$. The physician increasingly overuses AI by also including patients with $\alpha \in \left[\frac{\theta r_{AI}-c-(1-\rho)r_{AI}}{(\rho l'-l)\theta+(2\rho-1)r_{AI}}, \tilde{\alpha}'_1\right]$. However, for other patients, it either reduces or remains the same relative to the first patient-protection scheme. *Q.E.D.*

PROOF OF PROPOSITION 8. Consider the range of $\theta r_{AI} - c$ in Proposition A1 in which the physician uses AI for patients with $\alpha \in [0, \tilde{\alpha}'_1]$ or $\alpha \in [\tilde{\alpha}'_2, 1]$. Because $\tilde{\alpha}'_1 = \frac{(\theta r_{AI}-c)-(1-\rho)r_{AI}}{\theta(\rho l'-l)+(2\rho-1)r_{AI}}$ decreases in ρ and $\tilde{\alpha}'_2 = \frac{\theta\rho l'+\rho r_{AI}-(\theta r_{AI}-c)}{\theta\rho l'+(2\rho-1)r_{AI}}$ increases in ρ, as ρ increases, both regions ($[0, \tilde{\alpha}'_1]$ and $[\tilde{\alpha}'_2, 1]$) shrink, meaning the physician is less likely to use AI. *Q.E.D.*

Proofs of Results for Extensions

Patient Utility from AI-Augmented Physician

We consider the case in which the physician has AI capabilities. Recall from Proposition 1 that we have the following four cases:

(i) If $\theta r_{AI} - c < -\hat{\alpha}_0 \left[(2\rho - 1)(u + h + \theta l - \theta r) - (1 - \rho)\theta l_{AI}\right]$, the physician does not use AI for any patients. Thus, the patient's expected utility is the same as that expressed in (4). Thus, whether the physician has access to the AI system in making treatment-plan decisions makes no difference.

(ii) If $-\hat{\alpha}_0 \left[(2\rho-1)\left(u+h+\theta l-\theta r\right)-(1-\rho)\,\theta l_{AI}\right] \leq \theta r_{AI}-c<0$, the physician uses AI if and only if $\underline{\alpha} = \frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha < \bar{\alpha} = \frac{\rho(u+h+\theta r)+\theta r_{AI}-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}$. The patient's utility representation can take four possible forms (note $\underline{\alpha} < \hat{\alpha}_0 \leq \bar{\alpha}$ and $\alpha_L < \hat{\alpha}_0 < \alpha_H$ mean these are the only five possible scenarios):

a. If $\alpha_L \leq \underline{\alpha} < \alpha_H \leq \bar{\alpha}$, the physician uses AI and follows AI's recommendation if $\alpha = \alpha_H$ and does not use AI if $\alpha = \alpha_L$. Thus, the patient's expected utility is

$$\beta[\rho u-(1-\rho)h-c]+(1-\beta)[(1-\alpha_L)u-\alpha_L h]. \tag{A7}$$

By comparing (A7) with (4), we can show the patient prefers an AI-augmented physician if

$$\rho > \alpha_H + \frac{c}{u+h}$$

and prefers a non-AI-augmented physician otherwise.

b. If $\underline{\alpha} < \alpha_L < \alpha_H \leq \bar{\alpha}$, the physician always uses AI and follows AI's recommendation. Thus, the patient's expected utility is

$$\rho u-(1-\rho)h-c. \tag{A8}$$

By comparing (A8) with (4), we can show the patient prefers an AI-augmented physician if

$$\rho > \beta\alpha_H + (1-\beta)(1-\alpha_L) + \frac{c}{u+h}$$

and prefers a non-AI-augmented physician otherwise.

c. If $\underline{\alpha} < \alpha_L \leq \bar{\alpha} < \alpha_H$, the physician uses AI if $\alpha = \alpha_L$ and does not use AI if $\alpha = \alpha_H$. Thus, the patient's expected utility is

$$\beta[\alpha_H u-(1-\alpha_H)h]+(1-\beta)[\rho u-(1-\rho)h-c]. \tag{A9}$$

By comparing (A9) with (4), we can show the patient prefers an AI-augmented physician if

$$\rho > 1 - \alpha_L + \frac{c}{u+h}$$

and prefers a non-AI-augmented physician otherwise.

e. If $\alpha_L \leq \underset{\sim}{\alpha} < \bar{\alpha} \leq \alpha_H$, the physician does not use AI for any patient. Thus, the patient's expected utility is the same as that expressed in (4). The patient is indifferent between an AI-augmented and a non-AI-augmented physician.

(iii) If $0 \leq \theta r_{AI} - c < (2\rho-1)\theta l_{AI}\hat{\alpha}_1$, the physician uses AI if and only if $0 \leq \alpha < \frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}}$ or $\frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha \leq 1$. By noting $\alpha_H > \underset{\sim}{\alpha} = \frac{(1-\rho)(u+h+\theta r)-\theta r_{AI}+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})}$, we have three subcases:

a. If $\alpha_L \leq \frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}} \leq \underset{\sim}{\alpha} < \alpha_H \leq \bar{\alpha}$, the physician always uses AI. When $\alpha = \alpha_L$, the physician always chooses the non-standard treatment plan; when $\alpha = \alpha_H$, the physician follows the AI system's recommendation. Thus, the patient's expected utility is

$$\beta[\rho u - (1-\rho)h] + (1-\beta)[(1-\alpha_L)u - \alpha_L h] - c. \tag{A10}$$

By comparing (A10) with (4), we can show the patient prefers an AI-augmented physician if

$$\rho > \alpha_H + \frac{c}{\beta(u+h)}$$

and prefers a non-AI-augmented physician otherwise.

b. If $\frac{\theta r_{AI}-c}{(2\rho-1)\theta l_{AI}} < \alpha_L \leq \underset{\sim}{\alpha} < \alpha_H \leq \bar{\alpha}$, the physician uses AI only when $\alpha = \alpha_H$. When $\alpha = \alpha_H$, the physician follows the AI system's recommendation; when $\alpha = \alpha_L$, the physician always chooses the non-standard treatment plan. Thus, the patient's expected utility is

$$\beta[\rho u - (1-\rho)h - c] + (1-\beta)[(1-\alpha_L)u - \alpha_L h]. \tag{A11}$$

By comparing (A11) with (4), we can show the patient prefers an AI-augmented physician if

$$\rho > \alpha_H + \frac{c}{u+h}$$

and prefers a non-AI-augmented physician otherwise.

c. If $\frac{\theta r_{AI} - c}{(2\rho-1)\theta l_{AI}} < \underline{\alpha} < \alpha_L < \alpha_H \leq \bar{\alpha}$, the physician always uses AI and follows AI recommendation. Thus, the patient's expected utility is

$$\rho u - (1-\rho)h - c. \tag{A12}$$

By comparing (A12) with (4), we can show the patient prefers an AI-augmented physician if

$$\rho > \beta\alpha_H + (1-\beta)(1-\alpha_L) + \frac{c}{u+h}$$

and prefers a non-AI-augmented physician otherwise.

d. If $\frac{\theta r_{AI} - c}{(2\rho-1)\theta l_{AI}} < \underline{\alpha} < \alpha_L < \bar{\alpha} < \alpha_H$, the physician always uses AI. The physician follows the AI recommendation if $\alpha = \alpha_L$ and chooses the non-standard treatment plan if $\alpha = \alpha_H$. Thus, the patient's expected utility is

$$\beta[\alpha_H u - (1-\alpha_H)h] + (1-\beta)[\rho u - (1-\rho)h] - c. \tag{A13}$$

By comparing (A13) with (4), we can show the patient prefers an AI-augmented physician if

$$\rho > 1 - \alpha_L + \frac{c}{(1-\beta)(u+h)}$$

and prefers a non-AI-augmented physician otherwise.

(iv) If $\theta r_{AI} - c \geq (2\rho - 1)\theta l_{AI}\hat{\alpha}_1$, the physician always uses AI for all patients. However, whether the physician follows AI recommendation depends on the relative magnitude of α_L, α_H, $\underline{\alpha}$, and $\bar{\alpha}$. The results will be similar to case (ii).

The result follows from all the above cases.

Data Network Effect

The data network effect brings the marginal benefit η of using AI on a patient. As shown in equation (A14), the parameter η appears with c. All the analyses and proofs for this section can be simply produced by replacing c with $c - \eta$ in our main analysis. Therefore, we do not repeat the proofs here.

AI Precision at a Cost to the Physician

Consider the patient-protection scheme in which the AI signal reinforces the standard of care. If the insurance company reimburses all AI use, the patient's expected utility is same as the main model and the physician's payoff simplifies to

$$\pi_{AI}(\alpha) = \theta(r_{AI} - k\rho) - c + \begin{cases} (1-\alpha)u + \theta r - \alpha\left[h + \theta\left(l + (2\rho - 1)\, l_{AI}\right)\right] & \text{if } \alpha < \hat{\alpha}_1, \\ \rho u + \left[\alpha(1-\rho) + (1-\alpha)\rho\right]\theta r - (1-\rho)\left[h + \theta\alpha\left(l - l_{AI}\right)\right] & \text{if } \hat{\alpha}_1 \leq \alpha < \hat{\alpha}_2 \\ \alpha u - (1-\alpha)h & \text{if } \alpha \geq \hat{\alpha}_2. \end{cases} \tag{A14}$$

A comparison of the physician's payoff when using AI and when not using AI reveals the physician's decision on whether to use AI is as follows:

(i) If $\theta(r_{AI} - k\rho) - c < -\hat{\alpha}_0\left[(2\rho - 1)\left(u + h + \theta l - \theta r\right) + (1-\rho)\,\theta l_{AI}\right]$, the physician does not use AI for any patients.

(ii) If $-\hat{\alpha}_0\left[(2\rho-1)\left(u+h+\theta l-\theta r\right)+(1-\rho)\,\theta l_{AI}\right] \leq \theta(r_{AI}-k\rho)-c<0$, the physician uses AI if and only if $\frac{(1-\rho)(u+h+\theta r)-\theta(r_{AI}-k\rho)+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha < \frac{\rho(u+h+\theta r)+\theta(r_{AI}-k\rho)-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}$.

(iii) If $0 \leq \theta(r_{AI}-k\rho)-c<(2\rho-1)\theta l_{AI}\hat{\alpha}_1$, the physician uses AI if and only if $0 \leq \alpha < \frac{\theta(r_{AI}-k\rho)-c}{(2\rho-1)\theta l_{AI}}$ or $\frac{(1-\rho)(u+h+\theta r)-\theta(r_{AI}-k\rho)+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha \leq 1$.

(iv) If $\theta(r_{AI}-k\rho)-c \geq (2\rho-1)\theta l_{AI}\hat{\alpha}_1$, the physician always uses AI for all patients.

The implication is that $\underline{r}$ becomes larger and $\bar{r}$ smaller. The parameter space $\underline{r} \leq r \leq \bar{r}$ in which the physician does not underuse AI becomes smaller, whereas the parameter space in which the physician underuses AI expands. As a result, the overall AI underuse expands. It is straightforward that the results presented in Proposition 3 continue to hold.

Note, $\frac{(r_{AI}-k\rho)-c}{(2\rho-1)\theta l_{AI}}$ is decreasing in ρ. Therefore, if $0 \leq \theta(r_{AI}-k\rho)-c<(2\rho-1)\theta l_{AI}\hat{\alpha}_1$, for patients with $0 \leq \alpha < \frac{\theta(r_{AI}-k\rho)-c}{(2\rho-1)\theta l_{AI}}$, the physician is less likely to use AI as ρ increases.

Now, suppose the insurance company reimburses for the physician's AI use only when AI signal helps in decision making. Similar to the main model, when the insurance company rejects the physician's claim, the burden of paying r_{AI} to the physician falls on the patient. The patient's expected utility as well as the expressions for thresholds $\check{\alpha}_1$ and $\check{\alpha}_2$ are the same as in the main model (for the corresponding reimbursement scheme). A comparison of the physician's payoff when using AI and when not using AI reveals the physician's decision on whether to use AI is as follows:

If the insurance company reimburses r_{AI} only when the physician follows the AI signal to prescribe the treatment plan, the physician's AI-use decision is the following:

(i) If $\theta(r_{AI}-k\rho)-c<-\hat{\alpha}_0\left[(2\rho-1)\left(u+h+\theta l-\theta r\right)+(1-\rho)\,\theta l_{AI}\right]$, the physician does not use AI for any patients.

(ii) If $-\hat{\alpha}_0\left[(2\rho-1)\left(u+h+\theta l-\theta r\right)+(1-\rho)\,\theta l_{AI}\right] \leq \theta(r_{AI}-k\rho)-c<(1-\rho)\,r_{AI}$, the physician uses AI if and only if $\frac{(1-\rho)(u+h+\theta r)-\theta(r_{AI}-k\rho)+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha < \frac{\rho(u+h+\theta r)+\theta(r_{AI}-k\rho)-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}$.

(iii) If $(1-\rho)\,r_{AI} \leq \theta(r_{AI}-k\rho)-c<\max\left\{(1-\rho)\,r_{AI}+(2\rho-1)\left(r_{AI}+\theta l_{AI}\right)\check{\alpha}_1, \rho r_{AI}-(2\rho-1)\,r_{AI}\check{\alpha}_2\right.$

the physician uses AI if and only if $0 \leq \alpha < \frac{(\theta(r_{AI}-k\rho)-c)-(1-\rho)r_{AI}}{(r_{AI}+\theta l_{AI})(2\rho-1)}$ or $\frac{(1-\rho)(u+h+\theta r)-\theta(r_{AI}-k\rho)+c}{u+h-\theta(2\rho-1)r+\theta l_{AI}+\theta\rho(l-l_{AI})} \leq \alpha < \frac{\rho(u+h+\theta r)+\theta(r_{AI}-k\rho)-c}{u+h+\theta(2\rho-1)r+\theta(1-\rho)(l-l_{AI})}$ or $\frac{\rho r_{AI}-(\theta(r_{AI}-k\rho)-c)}{(2\rho-1)r_{AI}} \leq \alpha \leq 1$.

(iv) If $\theta(r_{AI} - k\rho) - c \geq \max\{(1-\rho)\, r_{AI} + (2\rho-1)\,(r_{AI} + \theta l_{AI})\,\check{\alpha}_1, \rho r_{AI} - (2\rho-1)\, r_{AI}\check{\alpha}_2\}$, the physician uses AI for all patients.

Similar to the reimbursement scheme in which the insurance company always reimburses for the physician's AI use, we find the overall AI underuse expands and the results presented in Proposition 6 continue to hold. Note, $\frac{\rho r_{AI}-(\theta(r_{AI}-k\rho)-c)}{(2\rho-1)r_{AI}}$ decreases in ρ, if $\theta(r_{AI} - k\rho) - c \geq \frac{r_{AI}}{2}$. Therefore, the result that the physician maybe less likely to use AI with increase in ρ continues to exist.

In equilibrium, the insurance company reimburses the physician r_{AI} only when she follows the AI signal.

The analysis for the second patient-protection scheme is similar.