Physicians as Persuaders: Evidence from Hospitals in China*

Jia Xiang†

Abstract

I estimate a Bayesian persuasion model to examine how financial incentives and asymmetric information shape physician-patient interactions. This approach offers new insights into the role of insurance. First, the model predicts that patients' coinsurance moderates physicians' responsiveness to increases in service fees. This prediction is supported by a difference-in-differences analysis using Chinese health insurance claims data with random variation in physicians' reimbursement and patients' coinsurance rates. Second, the model implies that lower coinsurance rates reduce both patient price elasticity and skepticism, increasing the likelihood of physicians misdirecting patients toward unnecessary treatments. Using structural model estimates, I show that for a diagnosis where surgical treatment is discretionary, nearly half of the patients who received surgery would not have done so were they fully informed. Such misdirection from physicians is greater when coinsurance rates decrease, highlighting a new inefficiency channel beyond moral hazard. I decompose the effect of lowering coinsurance into moral hazard and the novel greater misdirection effects. Counterfactual analysis shows that, while patients benefit from lower out-of-pocket costs, greater misdirection nearly offsets these welfare gains.

Keywords: asymmetric information, insurance, moral hazard, financial incentives, persuasion

*First version: November 2019. I am deeply grateful to my advisors, Paul Grieco, Daniel Grodzicki, and Mark Roberts, for their invaluable guidance and support throughout this project. I also thank the participants of seminars and conferences for their insightful feedback. I acknowledge the anonymous data providers from a local Healthcare Security Administration in China. The views expressed in this paper are solely mine and do not necessarily reflect those of the administration. All remaining errors are my own.

†Business Economics and Public Policy, Kelley School of Business, Indiana University, Bloomington, IN 47401, USA. Email: xiangjia@iu.edu.

1 Introduction

Asymmetric information is a key feature in expert-client relationships. In many cases, informed experts, such as auto mechanics, financial advisers, and physicians, face fundamental conflicts of interest as they advise clients on needed services while also profiting from those services. Ultimately, clients have the final say by assessing potentially biased advice of the expert alongside their own financial considerations. These joint decisions often involve substantial financial and health consequences. Despite the critical roles of asymmetric information and financial incentives in expert-client relationships, research on the interplay among demand-side incentives, supply-side incentives, and asymmetric information and how they jointly shape decision-making is surprisingly limited.

This paper examines the interplay of financial incentives and asymmetric information in expert-client relationships in the context of the medical treatment choices made by physicians and patients. This is an important setting to study because it involves high-stakes decisions that affect healthcare spending and patient quality of life, especially in a society facing rapidly growing health expenditures and an aging population. It addresses three questions theoretically and empirically. First, how does patient price sensitivity influence physician responsiveness to payment? Second, to what extent are treatment choices contrary to patients' best interests due to information asymmetry? Third, how does this distortion depend on each side's financial incentives?

To answer these questions, my central contribution is to apply the theory of Bayesian persuasion (Aumann et al., 1995; Kamenica and Gentzkow, 2011) to study the information transmission in physician-patient interactions under the influence of both patient and physician financial incentives. The model brings novel insights that insurance plays a more important role in healthcare spending and efficiency than traditional wisdom, focused on moral hazard or patient price sensitivity, have suggested (Newhouse et al., 1993). First, the model predicts that lower coinsurance rates amplify physicians' responsiveness to financial incentives, influencing healthcare spending more than patient price elasticities alone. I validate this prediction by leveraging the exogenous variation in physician reimbursement and patient coinsurance rates using health insurance claims data in China. Second, the model implies that lower coinsurance rates reduce patient skepticism, making them more susceptible to physicians' misdirection into unnecessary treatments. This greater misdirection channel suggests that welfare analysis cannot rely solely on demand responses to coinsurance, as is common in empirical work (Feldman and Dowd, 1991). Welfare outcomes also depend on the extent of supply-side misdirection and its responsiveness to coinsurance rates. Greater misdirection harms patient welfare and exacerbates inefficiencies beyond the traditional moral hazard effect. To quantify the extent of physicians' misdirection and disentangle it from the moral hazard effect, I estimate a Bayesian persuasion model. For neck arthritis, where surgical treatment is discretionary, I find that nearly half of the patients who chose surgery would not have done so were they as informed as their physicians. Such misdirection is greater for the more insured patients. Counterfactual analysis shows that while lower coinsurance rates reduce patients' medical expenses, they can harm patient welfare due to greater misdirection.

The Bayesian persuasion model explicitly captures how skeptical patients integrate partially altruistic physicians' recommendations into their decisions, and how physicians incorporate patients' decision making processes into recommendation strategy. It considers a diagnosis with two treatment options: a low-cost non-surgical treatment and a more expensive surgery. The physician, assumed to value both patient well-being and personal compensation, commits to a recommendation strategy that accounts for patient characteristics, out-of-pocket price, and physician remuneration. Crucially, the physician faces an incentive compatibility constraint (or obedience constraint), ensuring the patient perceives the recommended treatment as at least as good as the alternative based on updated beliefs. Intuitively, if a physician pushes a treatment too aggressively, the patient may become suspicious, update their beliefs less favorably, and potentially reject the recommendation. In equilibrium, the recommendation strategy does not perfectly align with a patient's best interests (e.g., it over-prescribes surgery), but it remains aligned enough for patients to follow, even when they are aware of the physician's financial bias. Consequently, some patients who would benefit more from nonsurgical treatment are misdirected into surgery. However, physicians' ability to misdirect is limited by their altruism and the need to convince skeptical patients.

The model delivers four novel insights that are otherwise not apparent. First, physicians are more responsive to service fee increases when coinsurance is lower, as lower coinsurance reduces patients' skepticism. This insight suggests that both demand and supply side incentives must be considered when formulating policies, even targeting only one side. Second, it predicts a hump-shaped relationship between the likelihood of surgery and the charge differential between surgical and non-surgical treatments. Surgery rates initially rise as the charge differential increases but decline beyond a certain threshold due to the binding obedience constraint and the need to maintain patient trust. Third, physicians' misdirection is greater when coinsurance decreases, because more insured patients tend to be less price-sensitive and skeptical, making them easier to persuade. Lastly, equipping patients with the same information as physicians prevents physicians from misdirecting patients, even though physicians still face financial incentives to do so. This insight suggests that, while physician financial incentives influence treatment decisions, better patient information is an effective counterbalance (Johnson and Rehavi, 2016).

I estimate the Bayesian persuasion model to quantify the misdirection due to asymmetric information and its interplay with financial incentives, using health insurance claims data from a city in China. I focus on cervical spondylosis, commonly known as neck arthritis. It is an ideal condition to study treatment decisions under alternative financial incentives because, except for those few patients who suffer from a severe case, the mode of treatment is discretionary (Rowland, 1992; Choi et al., 2016).

What makes the Chinese health insurance claims data well-suited is that one can observe independent and exogenous variations in physicians' remunerations and patients' coinsurance rates, key to disentangling their roles. A policy change increased physicians' charge differential between surgical and non-surgical treatments. Patients' insurance types, tied to their Hukou registration (urban or rural, inherited from the mother and difficult to change), are also plausibly exogenous. Using a difference-in-differences analysis, I show that, for neck arthritis, surgery was chosen nearly three times as often after the policy change, with no clear impact on health outcomes measured by the 30-day readmission rate. Moreover, the effect was 1.5 times larger for more insured patients, consistent with the prediction that coinsurance moderates physicians' response to fee increases. The significance of such effects implies that the data variation regarding physicians' remuneration and coinsurance rates is sufficient for generating precise estimates of financial sensitivity for each side in the structural model.

Additionally, I present reduced-form evidence supporting the model-predicted hump-shaped relationship between the probability of surgery and the charge differential between surgical and non-surgical treatments, using an instrumental variables estimator. To parameterize the model, I assume a quadratic form of the patient's out-of-pocket cost and the physician's remuneration to capture potential income effects on both sides. To address potential bias in estimating financial sensitivities due to unobservable patient characteristics correlated with the charge differential between surgical and non-surgical treatments, I use the policy change affecting the charge differential as an instrumental variable and apply a control function approach in estimating the structural model.

Model estimates suggest that for neck arthritis, a physician with average financial incentives trades off 1 dollar of her remuneration for 8 dollars of patient out-of-pocket cost, placing a weight of 11 percent on patient utility. In addition, the effect of asymmetric information is sizable: while only 8 percent of patients choose surgery, almost half of them would have chosen non-surgical treatment if fully informed. The degree of misdirection varies with physician financial incentives: removing physicians' financial incentives achieves the same treatment decisions as fully informing patients (while keeping physicians' financial incentives) does. Most importantly, I show that the degree of misdirection depends on patient coinsurance: a larger proportion of the more insured patients are misdirected.

I consider a counterfactual policy when a physician's compensation increases from its current level of 20 percent to 100 percent of the hospital revenue generated by the physician. I first use this exercise to show that while this policy does not change patients' out-of-pocket price, patients' coinsurance rate affects how physicians respond. In this counterfactual simulation, the surgery rate rises by 18 percent for the less insured while by 23 percent for the more insured. Consequently, patient welfare decreases by a larger extent for the more insured. Next, I benchmark my predictions against data from a 2016 policy change that introduced the same physician compensation adjustment in another region of China. The surgery rate for randomly selected departments increased by 21 percent (Gong et al., 2021). Thus, the Bayesian persuasion model produces realistic counterfactual predictions.

While the above counterfactual considers only changes to physician compensation, insurance coverage may also change simultaneously. I find that for the counterfactual increase in physician remuneration, concurrently adjusting patient coinsurance rates can result in an increase, decrease, or no change in the surgery rate. This underscores the need to design incentives on both the demand and supply sides simultaneously to achieve the intended outcomes.

Lastly, model estimates allow a measurement of the new effect of lowering coinsurance rate, i.e., physicians' greater misdirection. This measurement is critical for any welfare analysis of insurance policy, and it is only feasible when patients' decision-making process is explicitly modeled. I consider a policy that lowers patient coinsurance rates by half. A decomposition of the effect shows that both traditional moral hazard and greater misdirection based on cost-sharing contribute to a 10 percent rise in surgery rate. The misdirection effect in particular accounts for one-fifth of this surge. While patient welfare increases by 8 percent due to a lower out-of-pocket price, greater misdirection almost completely offsets the patient welfare gain, indicating that some patients are worse off due to surgeries they would not have opted for otherwise.

It is worth mentioning that modeling patient skepticism via the obedience constraint, in addition to allowing physicians to weigh patient utility, is necessary for generating new insights. A model without the obedience constraint inevitably overlooks the new channel of greater misdirection. In Appendix I, I estimate an alternative model without the obedience constraint and consider the same counterfactual policies. I show that this alternative model leads to significantly different patient welfare implications when altering patient coinsurance rates, yet at the same time fails to replicate the observed real-world data from the 2016 policy change regarding physician remuneration. This highlights the importance of incorporating the obedience constraint when modeling physician-patient interactions in my application to the Chinese health insurance claims data. Using my Bayesian model estimates, I further demonstrate that the obedience constraint is empirically relevant, with 73 percent of physicians facing binding constraints when recommending surgery.

My analysis contributes to studies on health care providers' financial incentives (e.g., Yip, 1998; Gruber et al., 1999; Clemens and Gottlieb, 2014; Eliason et al., 2018; Einav et al., 2018) and those on patients' price sensitivity (e.g., Manning et al., 1987), usually done separately. My contribution lies in focusing on the physician-patient interaction and separately identifying their roles in joint decision-making. Among the few empirical studies estimating both patient and provider sensitivity to financial incentives (Iizuka, 2007, 2012; Dickstein, 2011; Sacks, 2018; Hackmann and Pohl, 2018), this paper shows how financial incentives interact and highlight the interplay between asymmetric information and financial incentives. The Bayesian persuasion model further decomposes cost-sharing effects into price sensitivity and physicians' greater misdirection.

This paper also adds to the discussion on the role of insurance. Early research identifies expanded insurance coverage as a key driver of rising healthcare costs (Feldstein, 1977). However, the RAND Health Insurance Experiment estimates the elasticity of demand with respect to patient cost-sharing at around -0.2 (Aron-Dine et al., 2013), suggesting a more limited role for insurance. Recent literature explain why market-wide changes in insurance may have a larger impact than individual-level changes (Finkelstein, 2007; Geddes and Schnell, 2023). My findings add a new dimension based on physician-patient interaction, showing that even at the individual level, insurance generosity amplifies physicians' responsiveness to financial incentives, thereby influencing healthcare spending more than patient price elasticities alone would suggest. This insight also helps explain why supply-side payment policies have a particularly strong impact on care utilization in the U.S.[1] This is because generous insurance in the U.S. enables a more salient supply-side response to financial incentives. Moreover, I highlight a new inefficiency channel beyond the traditional moral hazard effect: greater misdirection resulting from lower coinsurance. With greater misdirection, the optimal coinsurance rate should be higher than when the trade-off considers only demand-side factors. I extend the work of Baicker et al. (2015), which incorporates both moral and behavioral hazards, by emphasizing the critical role of physician-patient interactions in insurance design.

[1] For example, the elasticity of care provision to physician payment rates is 1.5 in Clemens and Gottlieb (2014) and 0.7 in Cabral et al. (2021).

My model advances earlier theoretical work on "physician-induced demand." Some models conceptualize the cost of inducement as a disutility in physicians' preferences (Ellis and McGuire, 1986; McGuire and Pauly, 1991), while others view it as reputational damage, such as patients' refusal of current or future services (Dranove, 1988; Pauly, 2009). I incorporate both physicians' preferences and their reputational concerns, using an information design framework. I demonstrate that such modeling is necessary since the obedience constraint is empirically relevant even when physicians are altruistic. My approach closely aligns with the patient refusal framework in Dranove (1988). I extend this analysis by applying the theory to data, leveraging unique variations in financial incentives. This allows me to estimate physicians' altruism, quantify unnecessary treatments, assess the welfare consequences of information asymmetry, and capture the greater misdirection channel.

To my knowledge, this paper is the first to empirically apply the information design framework by estimating a Bayesian persuasion model. Vatter (2021) and Decker (2022) are two later papers that empirically apply the information design framework to scoring design. All other applications of Bayesian persuasion are theoretical.[2]

The remainder of the paper is organized as follows. Section 2 describes the Bayesian persuasion model of physician-patient interaction. Section 3 provides empirical setting and data description. Section 4 tests the model implications, providing direct evidence regarding physicians' and patients' financial incentives. Section 5 discusses model estimation. Section 6 presents estimation results and counterfactual analyses regarding changing financial incentives and information environment. Section 7 concludes.

2 A Model of Physician-Patient Interaction

I present a model in the spirit of Kamenica and Gentzkow (2011) to study information transmission between a physician and a patient, so that it is tractable to study the effect of asymmetric information and the interaction of the financial incentives empirically.

2.1 The Treatment Decision Process

I consider a physician and a patient i with one diagnosis, for which two treatment options, surgical treatment (s) or drug treatment (d), are available. I normalize patient i's utility from d to 0. The relative value of surgery for patient i is

[2] See Kamenica (2019) for a list of theoretical applications.

$$
\begin{aligned}
U_i^s &= \xi_i - \kappa_i, \\
\text{where } \kappa_i &\equiv k\big(\underbrace{y_i \cdot \Delta T_i}_{\substack{\text{relative} \\ \text{out-of-pocket price}}}, \underbrace{\mathbf{Z}_i}_{\text{demographics}}\big), \\
\text{and } \Delta T_i &\equiv T_i^s - T_i^d.
\end{aligned} \tag{1}
$$

ξ_i represents the relative value of surgery, for example, the extent of pain relief after the treatment, which is unknown to either party before the patient walks in. The patient and physician share a common i.i.d prior for the true state of ξ_i, denoted by its cumulative distribution function, $G(\xi)$. However, different from the patient, the physician has the expertise to learn the exact value ξ_i through performing lab tests and analyzing the results. In other words, it is the ability to learn ξ_i that gives the physician her information advantage.

κ_i summarizes the common knowledge regarding the relative value of s treatment. It is assumed to be a function of i's demographics and the out-of-pocket price difference between surgical and drug treatment, which represents the patient's financial incentives.[3] y_i is patient coinsurance rate, and ΔT_i equals the hospital charge differential between treatment s and d, $T_i^s - T_i^d$. The relative out-of-pocket price does not depend on deductible or potential lump-sum subsidies because even the lowest charge for the analyzed diagnosis exceeds the highest possible deductible and subsidies.[4] The greater the κ_i, the less desirable the surgery is. It is reasonable to assume that $\frac{\partial \kappa_i}{\partial (y_i \cdot \Delta T_i)} > 0$, implying that the patient suffers from a higher out-of-pocket price. Were the patient fully informed, he would take surgery if and only if $\xi_i \geq \kappa_i$.

A physician's payoff from treatment d is normalized to be 0, so that the relative payoff from surgery, M_i^s, is:

$$
\begin{aligned}
M_i^s &= U_i^s + F_i, \\
\text{where } F_i &\equiv f(\Delta T_i).
\end{aligned} \tag{2}
$$

F_i represents a physician's utility from the relative remuneration of surgery. Since, in China, physicians' remuneration bonus is a fixed proportion of hospital revenue, I use $\Delta T_i \equiv T_i^s - T_i^d$ to

[3] In my empirical context, physicians observe insurance types from inpatient paperwork.

[4] A subsidy is a lump-sum deduction of the out-of-pocket cost. The eligibility and amount of the subsidy depend on whether the patient's annual income is below some threshold (for example, 250 percent of the local poverty line for a particular amount of subsidy), level of disability, veteran status, etc. The exact amount is reviewed and calculated by the government in an unknown way. However, the observed amount in the data ranges from several dollars to about one hundred dollars per visit.

represent the physician's financial incentives to perform a surgery.[5] It is reasonable to assume a physician gets positive utility from more remuneration, i.e., $f'(\Delta T_i) > 0$. A physician's payoff is a weighted sum of her patient's utility (U_i^s) and her own financial incentive (ΔT_i). The parameters in $f(\Delta T_i)$ govern the weight. As F_i decreases, the incentives are more aligned. The perfect incentive alignment occurs at $F_i = 0$. I assume that except for ξ_i, the patient observes the physician's preferences, including the physician's financial incentives.[6]

Importantly, the charge differential, ΔT_i, measures the conflict of interest because it constitutes part of physicians' remuneration as well as costs to patients. As ΔT_i increases, physicians are more incentivized towards surgery, while the patients are less incentivized.

In the game: (1) the physician moves first and commits to her strategy. She designs a treatment recommendation strategy $\sigma(\xi, \kappa, F) : \mathbb{R}^3 \to \Delta(\{s, d\})$, a mapping from triples (ξ, κ, F) to the set of all probability distributions over $\{s, d\}$.[7] This strategy essentially controls the accuracy of information provided to the patient on ξ_i. The recommendation will depend on patient i's out-of-pocket price and demographics, as well as on the physician's financial incentives; (2) patient i arrives, and the state ξ_i is realized. The patient then receives a recommendation, $r_i \in \{s, d\}$, according to the physician's committed strategy; (3) the patient updates his belief on ξ_i based on his prior and physician's recommendation; (4) the patient decides on $\Delta(\{s, d\})$.

The solution concept is physician-preferred subgame perfect equilibrium: if the patient is indifferent between surgery and drugs at a given belief, he is assumed to take the action that maximizes the physician's expected utility.

2.2 Patient's Problem and Physician's Problem

Patient's Problem. The patient takes as given the recommendation $r_i \in \{s, d\}$ and the recommendation strategy $\sigma_i(\xi, \kappa, F) = \{p_i^s, p_i^d\}$, which specifies the probabilities of recommending each option. Given a surgery recommendation, the patient updates his expected utility from surgery, using Bayes' rule:

[5] Physicians' bonuses are usually 20 percent of the revenue they create. This percentage will be factored in when calculating physicians' sensitivity to financial incentives in Section 6.1. To the extent that physicians' costs, such as efforts and malpractice risk, could differ between the two treatment options, Appendix F.2 assesses these costs with supplemental evidence and concludes that the cost differences are limited in this setting.

[6] My survey of 58 patients in China confirms that patients know about physicians' payment schemes from anecdotes. The prevalent practice of informal payments for medical services in China also reflects patients' concerns about physician agency problems (Li et al., 2022). In terms of their awareness of physicians' exact financial incentives, patients and their families can check the regulated price of each procedure on rolling screens in hospital lobbies. Moreover, physicians in China can typically provide a rough estimate of a patient's medical expenditure for various treatments, as pricing is standardized by the government and consistent across different insurance types.

[7] Without loss of generality, I restrict the physician's message space to be the action space, i.e., the set of treatment modes.

$$\mathbb{E}\big(U_i^s|r_i = s, \sigma_i(\cdot)\big) = \mathbb{E}\big(U_i^s\big|\xi_i \in A_i\big) = \mathbb{E}\big(\xi_i\big|\xi_i \in A_i\big) - \kappa_i, \tag{3}$$

where $A_i = \{\xi : p_i^s > 0\}$, the set of values for ξ that induces the surgery recommendation to i with positive probability. If the physician recommends d, the patient updates

$$\mathbb{E}\big(U_i^s|r_i = d, \sigma_i(\cdot)\big) = \mathbb{E}\big(\xi_i\big|\xi_i \in B_i\big) - \kappa_i, \tag{4}$$

where $B_i = \{\xi : p_i^d > 0\}$, the set of ξ that induces the recommendation of d to i with positive probability. Therefore, given $r_i = s$, the patient's best response is to follow the recommendation (s) if and only if $\mathbb{E}\big(U_i^s|r_i = s, \sigma_i(\cdot)\big) \geq 0$. Given $r_i = d$, it is optimal to follow the recommendation (d) if and only if $\mathbb{E}\big(U_i^s|r_i = d, \sigma_i(\cdot)\big) \leq 0$.

Physician's Problem. I focus on the obedient recommendation strategy, as in Kamenica and Gentzkow (2011) and Bergemann and Morris (2019). That is, before the realization of ξ_i, and given patient i's best response to the recommendation, the physician chooses a recommendation strategy $\sigma_i(\cdot)$ to maximize her expected payoff, such that the Bayesian patient is willing to follow her recommendation even though he knows it is biased:

$$\begin{aligned} &\max_{\sigma_i(\cdot)} \int_{\xi \in A_i} p_i^s \cdot \underbrace{\big(\xi - \kappa_i + F_i\big)}_{\text{utility if } s} dG(\xi) + \int_{\xi \in B_i} p_i^d \cdot \underbrace{0}_{\text{utility if } d} dG(\xi) \\ &\textit{s.t (Obedience constraints)} \\ &\quad ① \ \mathbb{E}\big(\xi_i\big|\xi_i \in A_i\big) - \kappa_i \geq 0 \\ &\quad ② \ \mathbb{E}\big(\xi_i\big|\xi_i \in B_i\big) - \kappa_i \leq 0. \end{aligned} \tag{5}$$

Physician's Cutoff Strategy. Since the physician's payoff is monotonic in ξ_i, the physician's optimal strategy takes the form

$$p_i^s = \begin{cases} 1 \ \text{ if } \ \xi_i \geq c_i \\ 0 \ \text{ otherwise,} \end{cases} \tag{6}$$

where c_i is a cutoff for ξ_i, above which she recommends surgery with probability 1, and drugs otherwise.[8] Appendix A.1 provides the arguments. The intuition is as follows: if the physician optimally recommends drugs with positive probability for some region of ξ_i above c_i, then there must exist $\tilde{c}_i > c_i$, and the physician recommends surgery if and only if $\xi_i > \tilde{c}_i$, such that the

[8] Strictly speaking, because each point has measure zero, physician's optimal strategy should be a cutoff strategy almost everywhere (up to a countable collection of points).

overall probability of recommending surgery remains the same. However, the physician has a higher expected utility due to the higher expected ξ_i. Cutoff $\tilde{c}_i$ then constitutes a profitable deviation.

With the cutoff strategy, it is straightforward to express the update in the patient's beliefs. For example, given $r_i = s$,

$$\mathbb{E}(U_i^s | r_i = s, c_i) = \mathbb{E}(\xi_i | \xi_i \geq c_i) - \kappa_i = \frac{\int_{c_i}^{\infty} \xi \cdot g(\xi) d\xi}{1 - G(c_i)} - \kappa_i, \tag{7}$$

where $g(\cdot)$ is the pdf of ξ_i.

Then the physician's optimization problem is reformulated as:

$$\begin{aligned} & \max_{c_i} \quad \int_{c_i}^{\infty} \big(\xi - \kappa_i + F_i\big)\, dG(\xi) + \int_{-\infty}^{c_i} 0 \, dG(\xi) \\ \iff \quad & \max_{c_i} \quad \big[1 - G(c_i)\big] \cdot \big[\mathbb{E}(\xi_i | \xi_i \geq c_i) - \kappa_i + F_i\big] \\ & \textit{s.t (Obedience constraints)} \\ & \quad ① \ \mathbb{E}(\xi_i | \xi_i \geq c_i) - \kappa_i \geq 0 \\ & \quad ② \ \mathbb{E}(\xi_i | \xi_i < c_i) - \kappa_i \leq 0. \end{aligned} \tag{8}$$

This formulation indicates that a physician's misdirection is constrained in two ways. First, she is altruistic: according to (8), even though lowering the cutoff increases the probability of surgery (increases $1-G(c_i)$) and brings more expected remuneration, it also leads a patient with a lower value of surgery to choose surgery (decreases $\mathbb{E}(\xi_i|\xi_i \geq c_i)$), which hurts the physician. Second, the physician's bias is restrained by the patient's rational judgment of the recommendation, captured by the obedience constraints. The conditional expectations in the constraints imply that the patient is aware of the physician's biased recommendation strategy. If the physician always recommends a particular type of treatment, the patient's updated expectation will coincide with his prior. In this case, the recommendation is uninformative and cannot influence the patient's decision.

2.3 Equilibrium

Since the patient's incentive compatibility is built into the physician's obedience constraints, the recommendation will produce the patient's equilibrium action.

I solve for the physician's obedient recommendation strategy using Karush-Kuhn -Tucker conditions in Appendix A.2. Under the assumption that physicians can always earn more from surgery ($F_i \geq 0$), which is empirically true for the diagnosis on which I focus, the second obedience constraint will not bind, and the optimal cutoff c_i^* can be written as

$$c_i^* = \max\{h^{-1}(\kappa_i),\ \ \kappa_i - F_i\}, \tag{9}$$

where $h^{-1}(x)$ is the inverse function of $h(x) = \mathbb{E}(\xi_i|\xi_i \geq x) \equiv \frac{\int_x^\infty \xi g(\xi) d\xi}{1-G(x)}$.[9]

$c_i^* = h^{-1}(\kappa_i)$ corresponds to the corner solution when constraint ① binds.[10] Under this optimal signal, the patient is exactly indifferent between s and d when being recommended surgery. $c_i^* = \kappa_i - F_i$ is the interior solution where the constraint does not bind.

2.4 Model Implications

I highlight four properties from the model. Appendix A.3 provides proof.

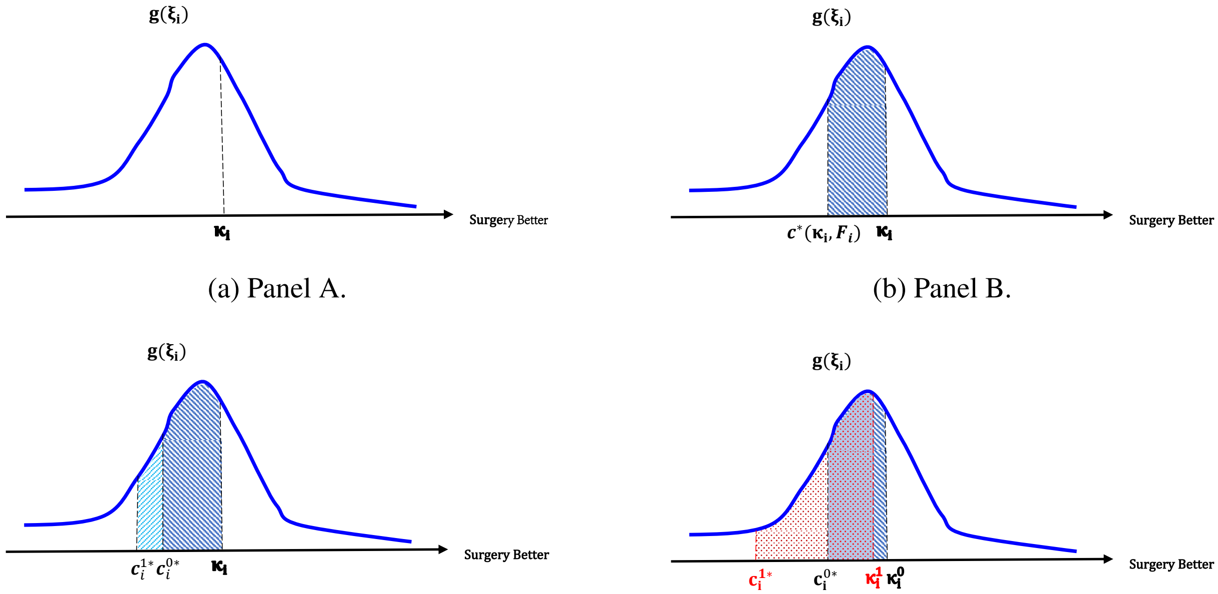
First of all, note that without any physician financial incentives ($F_i = 0$), the perfect agent would choose $c_i^* = \kappa_i$. I refer to it as the first-best cutoff from the patient's perspective, as illustrated in Panel A of Figure 1.

Property 1. *The magnitude of a physician's response to financial incentives,* $|\frac{\partial c_i^*}{\partial F_i}|$*, decreases in* κ_i*. That is, when patients are more insured and thus less resistant to surgery (i.e.,* κ_i *smaller), physicians become more responsive to their own financial incentives,* F_i*.*

This property concerns the cross partial $\frac{\partial}{\partial \kappa_i}\left(\frac{\partial c_i^*}{\partial F_i}\right)$. When physicians' financial incentives F_i increase, such as through higher surgery fees, if patients are well insured, physicians can respond by lowering the cutoff to the left of κ_i without worrying about patients' refusal. Graphically, comparing Panel B to Panel A of Figure 1, some patients who are better candidates for drug treatment (the shaded area) are pooled with those who indeed benefit from surgery (to the right of the shaded area). The shaded area represents the probability (portion) of patients being persuaded to opt for surgery due to inaccurate information. The model also implies that, due to the physician's altruism, she only misdirects marginal patients — those who would be only slightly worse off with surgery. However, if patients are less insured, they tend to be more skeptical. As a result, the obedience constraint is more likely to bind, limiting physicians' ability to respond to any fee increase.

[9] The expression for optimal cutoff holds for any $F_i > 0$. For $F_i = 0$, it holds for all prior distributions with thin right tails (see Appendix A.2). The interpretation of a thin right tail is that it is rare to find a patient whose value of surgery is extremely high. For a diagnosis with a very low surgery rate, such as the neck arthritis that I focus on, this is a reasonable assumption. Therefore, I adopt this assumption to make the expression of the optimal cutoff concise.

[10] I refer to constraint ① as the "obedience constraint" hereafter.



(a) Panel A.

(b) Panel B.

(c) Panel C.

(d) Panel D.

Figure 1: Physician's Optimal Cutoff Under Different Financial Incentives

Note: *This figure illustrates how a physician's optimal cutoff for her privately known benefit of surgery, ξ_i, responds to an increase in her financial incentives (Panels A, B, and C) or a decrease in the patient's coinsurance rate (Panel D). The shapes are illustrative.*

Property 2. *c_i^* is strictly decreasing in charge differential ΔT_i, only if the obedience constraint does not bind and physician incentives dominate, and is increasing in ΔT_i once the constraint binds. That is, the probability of surgery strictly increases with charge differential ΔT_i for the unconstrained physicians, but decreases for the constrained physicians.*

Property 2 is about $\frac{\partial(c_i^*)}{\partial(\Delta T_i)}$. It describes a non-monotonic relationship between the probability of surgery and the level of conflict of interest, measured by ΔT_i. When the conflict of interest is mild, the obedience constraint does not bind, and both physician and patient incentives are at play. If physicians' incentives dominate ($f'(\Delta T_i) > \frac{\partial k_i}{\partial \Delta T_i}$, i.e., physicians are more sensitive to the charge differential), increasing the relative remuneration from surgery motivates physicians to lower the cutoff to the left of κ_i without concerns for patients' refusal. In this case, the probability of surgery rises as the charge differential widens.

Once the charge differential, or conflict of interest, ΔT_i reaches a certain threshold, the obedience constraint binds, making physician incentives irrelevant. At this point, constrained physicians optimally set the cutoff so that patients remain exactly indifferent between treatment options. Consequently, to maintain this indifference and patient trust, the probability of surgery must decline as ΔT_i rises. Graphically, in Panel C of Figure 1, as ΔT_i continues to increase, the cutoff initially shifts further from the first best cutoff κ_i until reaching the threshold c_i^{1*}, after which it moves closer to the first best κ_i.

Define misdirection as the distance between the physician's cutoff and the first best cutoff. The change in misdirection with respect to the conflict of interest is

$$\frac{\partial(\kappa_i - c_i^*)}{\partial(\Delta T_i)} = \mathbb{1}\{h^{-1}(\kappa_i) < \kappa_i - F_i\} \cdot \underbrace{f^{'}(\Delta T_i)}_{>0} + \mathbb{1}\{h^{-1}(\kappa_i) \geq \kappa_i - F_i\} \cdot \underbrace{\frac{\partial k_i}{\partial \Delta T_i} \cdot (1 - \frac{1}{h^{'}(\kappa_i)})}_{<0 \text{ by Appendix (A.22)}}. \tag{10}$$

Therefore, an unconstrained physician increases misdirection in response to a rise in ΔT_i, while a constrained physician decreases misdirection. Similar to Property 1, there is an analogous implication regarding how patients' financial incentives, included in κ_i, influence physicians' responses to charge differentials, $\frac{\partial(\kappa_i - c_i^*)}{\partial(\Delta T_i)}$. When patients are more insured, physicians are not constrained and can misdirect more patients in response to a large charge differential. Conversely, when patients have less insurance coverage, physicians are likely constrained and must misdirect fewer patients as the charge differential increases.

Having discussed the interaction between physician and patient incentives, I now use the following two properties to examine the interplay between asymmetric information and each party's incentives.

Property 3. *Holding physicians' financial incentives fixed, a physician's misdirection due to asymmetric information decreases in κ_i, when the obedience constraint is binding. That is, the less resistant the patient is to surgery (for example, the more insured), the more likely the patient is to be persuaded by the physician.*

This property is about $\frac{\partial(\kappa_i - c_i^*)}{\partial \kappa_i}$. Intuitively, as patients become less price sensitive, the physician can use a strategy that is more likely to recommend surgery given patient obedience and identical health conditions. Graphically, Panel D of Figure 1 suggests that when a patient is more insured (smaller y_i and k_i), his first best cutoff will be lower because he is less sensitive to price. This moral hazard effect is the cutoff movement from κ_i^0 to κ_i^1. In addition to this effect, the physician can now further lower her cutoff to c_i^{1*}, since the more-insured patient is more likely to agree with the recommendation. Ultimately, the new misdirected portion under a more generous insurance, $\kappa_i^1 - c_i^{1*}$, is greater than $\kappa_i^0 - c_i^{0*}$.

Formally, for a constrained physician, her initial cutoff is $c_i^{0*} = h^{-1}(\kappa_i^0)$. When the patient's coinsurance rate decreases, the cutoff movement is decomposed into two parts: one due to moral hazard and the other arising from greater misdirection.

$$\underbrace{\left|h^{-1}(\kappa_i^1) - h^{-1}(\kappa_i^0)\right|}_{\text{Total effect of lowering coinsurance}} = h^{-1}(\kappa_i^0) - h^{-1}(\kappa_i^1) = \underbrace{\kappa_i^0 - \kappa_i^1}_{\text{moral hazard}} + \underbrace{\left(\kappa_i^1 - h^{-1}(\kappa_i^1)\right) - \left(\kappa_i^0 - h^{-1}(\kappa_i^0)\right)}_{\substack{\text{greater misdirection} \\ > 0 \text{ by Appendix (A.22)}}}. \quad (11)$$

The greater misdirection channel is essential for any welfare analysis for coinsurance rate adjustments, particularly in terms of total inefficiency and patient welfare. Figure 2 illustrates the welfare impact of a coinsurance rate decrease, accounting for both demand response and physician misdirection. The utilization curve reflects how coinsurance rate affects care demand, which is influenced by patient price elasticity and physician misdirection. Due to physicians' misdirection based on coinsurance, this demand curve neither equates to nor parallels the patient's marginal private health benefit.

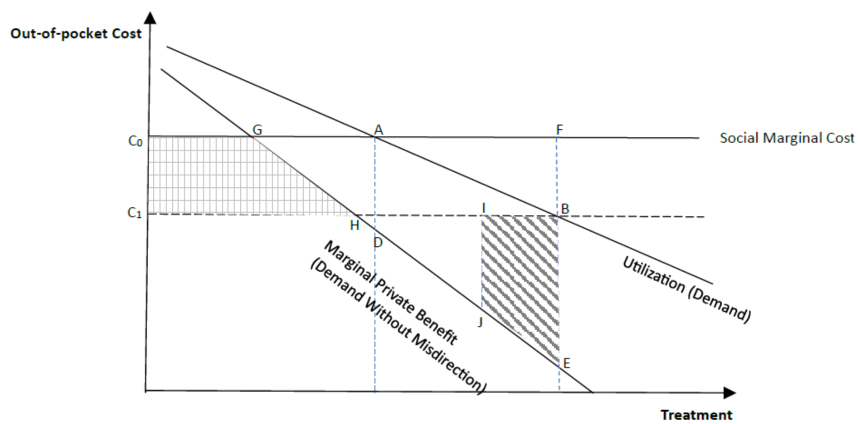


Figure 2: Welfare Impact of a Coinsurance Rate Decrease: Misdirection versus Moral Hazard

Note: *This figure illustrates that when the coinsurance rate decreases and out-of-pocket costs drop from C_0 to C_1, moral hazard leads to a patient welfare gain represented by the area C_0GHC_1, while greater misdirection results in a patient welfare loss of $IBEJ$.*

When the out-of-pocket cost equals society's cost, C_0, the gap between the marginal private benefit and utilization reflects misdirection, represented as GA. This misdirection leads to a patient welfare loss equal to the area GAD, as patients undergoing surgery solely due to physician misdirection have private benefits below its cost. When the coinsurance rate decreases, reducing the out-of-pocket cost to C_1, misdirection increases to HB, which is IB greater than the initial misdirection (assuming $GA = HI$). As a result, patient welfare loss from misdirection rises to HBE, with the greater welfare loss represented by the trapezoid area $IBEJ$. Pure moral hazard, on the other hand, represents the movement from G to H, increasing patient welfare by the area C_0GHC_1. The net patient welfare change depends on the trade-off between the welfare loss from misdirection ($IBEJ$) and the welfare gain from cheaper medical bill (C_0GHC_1). In my empirical application, these opposing effects approximately cancel each other out, resulting in a near-zero net patient welfare effect.

The total welfare loss is AFB if we rely solely on demand response as the sufficient statistic for welfare, as is typically done.[11] When misdirection is considered, the efficiency loss increases and is represented by $AFED$. Furthermore, compared to the scenario where the demand and marginal benefit curves are parallel, greater misdirection driven by coinsurance further exacerbates the inefficiency loss.

This model property suggests that the optimal coinsurance rate should be higher when accounting for greater misdirection, compared to when the trade-off considers only demand-side factors.

Property 4. *Providing a patient with full information forces the physician to give recommendations based solely on the patient's utility ($c_i^* = \kappa_i$), regardless of the physician's financial incentives. The patient achieves the full information payoff if the physician has no financial incentives ($F_i = 0$).*

Property 4 suggests two substitutable approaches to limit physicians' propensity to recommend unnecessary surgeries: reducing information asymmetry or minimizing the influence of physicians' financial incentives. In other words, while financial incentives can shape treatment decisions, patient information or skepticism serves as a powerful counterbalance (Johnson and Rehavi, 2016).

2.5 Discussion

Connection to Kamenica and Gentzkow (2011) (hereafter KG). The modeling of physician-patient interaction in surgical decisions differs from the canonical example of a prosecutor persuading a judge in KG in two ways. First, the state variable in my case, κ_i, is continuous. I therefore prove that a cutoff strategy is optimal in this setting. Second, the physician, who acts as the message sender, is altruistic, resulting in an equilibrium where the obedience constraint does not necessarily bind. Appendix A.4 further highlights the connections and distinctions with KG.

Commitment Assumption. The assumption of physicians' commitment to the recommendation rule can be relaxed. For example, Lin and Liu (2022) shows that, in my case, it is sufficient to assume patients observe the distribution of recommendations. This is likely, given

[11] An exception is Baicker et al. (2015), which distinguishes the demand curve from the marginal private benefit curve by accounting for patients making mistakes. While I abstract away such behavioral hazard here, incorporating this layer, where demand without misdirection deviates from the marginal private benefit, would not undermine the intuition that greater misdirection is welfare relevant.

that some hospitals provide transparency dashboards or decision aids that display statistics on treatment options recommended for patients with similar conditions. Alternatively, Nguyen and Tan (2021) considers a lying cost that can help achieve the full commitment payoff without assuming full commitment. In addition, the model is chosen over other games of communication because it guarantees the uniqueness of equilibrium so that the model predictions are sharp.[12]

Alternative Models. Often times, empirical models of physician behavior involve a single decision-maker considering a weighted sum of the patient's utility and the physician's utility (Chandra et al., 2011; Currie and MacLeod, 2017). The unique part of the persuasion model is the explicitly active role of the patient's skepticism and belief updating in the treatment decision, modeled through the obedience constraint. Appendix I shows that the obedience constraint is crucial to generate the model implications regarding how patient and physician financial incentives interact and their interplay with asymmetric information.

3 Empirical Setting

In this section, I describe the variation in physicians' remuneration differential between surgical and non-surgical treatments, as well as in patients' coinsurance rates.

3.1 Physician Income and Patient Insurance

My empirical analysis is based on a city of average size in China. Hospitals in this area are paid on a fee-for-service basis, with the fees set by the government. Under fee-for-service, to treat neck arthritis, for example, a hospital will bill for the physician's consultation, each test (e.g., X-Rays, CT scan, MRI, or blood tests), each procedure (acupuncture, physical therapy, or surgery), and daily inpatient care (bed and nursing) — all independently. This payment scheme is in contrast to the bundled payment scheme, whereby a hospital is paid a single price for all services needed for an entire episode of care, such as the diagnosis-related group system adopted by the Centers for Medicare and Medicaid Services in the U.S.

Physicians work as employees of hospitals. Their income consists of two parts: a fixed salary, and a variable component, which is a commission percentage of the patient care revenues

[12] In a cheap-talk model, if all individuals are assumed to play the same equilibrium, either the babbling or the influential one (whenever possible), the estimated model cannot predict the hump-shape of surgery probability. It could be that, under some convexification for the equilibria, the model predicts the surgery patterns. As future work, it would be interesting to derive the bound of estimates in a cheap-talk setting. Nonetheless, in my survey of 11 orthopedic physicians and 58 patients in China suggests that, over 95 percent of the time, patients follow physicians' advice. This survey result favors a model with a unique equilibrium.

that the physician generates.[13] The fixed salary accounts for less than a quarter of the physician's total income (Ran et al., 2013).

In China, there are three types of medical insurance: government-provided urban employment insurance, government-provided resident insurance, and private insurance. The two government-provided insurances are exclusive, and the private insurance can be purchased either in addition to the government-provided or on its own. I do not observe any private insurance claims information; however, excluding private claims still yields a broadly representative sample, as only about 1 in 20 people in China own private insurance.[14] Thus, I focus on government-provided insurance alone, which covers over 95 percent of the study population.

An individual is eligible and mandated to take urban employment insurance if he works in a formal sector and holds an urban registration ("hukou" in Chinese). All other citizens — those with rural registration or those with urban registration but without a formal-sector job — are eligible for the resident insurance. It is reasonable to assume that people do not self-select into an insurance type, because the urban-rural registration is given at birth and is extremely hard to convert from one type to another.[15] To the extent that individuals registered as urban could choose to be formally employed to be eligible for the more generous insurance, I argue that insurance type is not associated with observed patient characteristics in my data or a large set of variables determining medical choices using external data in Section 4.2.

Conditional on hospital tier in a given year, the beneficiaries within the same insurance type get the same coverage in terms of deductible, coinsurance rate, and reimbursement ceiling. Appendix B includes a generosity comparison between the two government-provided insurance types. Overall, the urban employment insurance provides a lower marginal out-of-pocket price and thus is more generous, while both insurance types face the same total prices set by the government. In the following analysis, I will refer to the patients with urban employment insurance as the *more insured* and the patients with resident insurance as the *less insured*.

Under either insurance, the total charge for a neck arthritis patient always exceeds the deductible and would not reach the annual coverage limit just because of this disease alone. Therefore, the out-of-pocket price difference by insurance type is primarily due to the difference in coinsurance rates: the averaged coinsurance rate across hospital tiers and plan-years is 0.18 for

[13]The commission percentage is typically 15–20 percent for hospitals in China Gong et al. (2021). Conversations with physicians in my study city confirm this.

[14]http://image-src.bcg.com/Images/BCG-Opportunities-Open-Up-Chinese-Private-Health-Insurance-Aug-$2016_tcm30-63204.pdf$

[15]The "hukou" categorizes each Chinese citizen as either rural or urban by place of origin. A person's "hukou" is usually inherited from his mother. The conversion from rural to urban can only happen through recruitment by a state-owned enterprise, enrollment in a higher education institution, promotion to a senior administrative job, and migration for personal reasons. Moreover, there is an annual quota for the conversion, fixed at about 0.15 percent of the urban population in each locale (Chan and Zhang, 1999).

the more insured and 0.28 for the less insured. Although variation in coinsurance exists within insurance type, once conditional on the hospital tier and plan-year, the variation in coinsurance rate that identifies patient price sensitivity only comes from insurance types.

3.2 Policy Change: Regulation and Drug Mark-up Cancellation

Public hospitals receive funding through service charges, drug sales, and government budget allocation. Since drugs for inpatient care are always filled in hospital-owned pharmacies, drug sales also contribute to hospital revenue.[16] The Chinese government regulates the prices of both medical services and drugs in all public hospitals.[17] Historically, the government strictly controlled the price to make healthcare more affordable. However, in the 1980s, the government allowed hospitals to mark up drugs by up to 15 percent above procurement prices, as a means of offsetting some hospital revenue lost through diminished government funding.[18] In fact, the 15 percent mark-up restriction was usually binding.

Although this 15 percent mark-up policy helped fund public hospitals' operations, "gradually it evolved into a way to reap profits, contributing to worsening problems such as over-prescribing, excessive use of antibiotics by hospitals, and rising medical expenses." according to Hesheng Wang, vice-minister of the Health and Family Planning Agency.[19] Partly due to this over-prescription, China's health care expenditure has outpaced its GDP growth in the past two decades.

In April 2009, the central government, through the *Suggestions of the CPC Central Committee and the State Council on Deepening the Reform of the Health Care System*, stated that the drug mark-up should be gradually eliminated in public hospitals. The cancellation of the drug mark-up forced hospitals to sell drugs at their purchase cost, which was uniform across hospitals due to centralized procurement bidding at the provincial level. In order to compensate partially for hospitals' revenue loss, meanwhile, the government raised regulated prices for medical services, requiring that raised prices for medical services must compensate for more than 80 percent of the potential loss of drug revenue. The central government set the goals and macro strategies that served as the basis for provincial implementation. Provincial governments

[16] Although the drugs for outpatient care can be filled in drug stores outside the hospitals, usually, patients' insurance would not reimburse the expenditures in those stores, making this option much less attractive.

[17] After 2015, private hospitals were allowed to set their market price. Due to the monitoring by governmental price regulation agencies and the competition with public hospitals, the price set by private hospitals was very close to the regulated price.

[18] Government funding decreased from about 60 percent of the total hospitals' income in the 1970s to less than 10 percent in the 2010s, according to the Ministry of Health's public records.

[19] Wang, Xiaodong. 2017. "Public hospitals told to end drug markups." China Daily, April 13. https://www.chinadaily.com.cn/china/2017-04/13/content$_2$8904319.htm.

were responsible for setting the operational details, such as determining the implementation schedule and adjusting service prices (Fu et al., 2018).

I refer to the cancellation of the drug mark-up and the adjustments of procedure prices as *the price menu change* or *the policy change*. The policy's introduction was staggered in the order of primary, county-level, and city-level hospitals, its roll-out decided by the central government. Country-wide implementation entailed three phases.

In the first phase, by April 2012, all the primary healthcare institutions, including village clinics, township health centers, and community health centers, across the country implemented the policy. In the second phase, by the end of 2015, all county-level hospitals were obligated to have implemented it. In the third phase, the policy reached all city-level hospitals by September 30, 2017. By this time, the policy had taken effect nationwide.

Study Region. I study a region in a western Chinese city. The sample region, featuring almost one million population, is the central area of the city. The per capita annual income in 2016 is about $ 6,600, which is comparable to the national average (China Statistical Yearbook 2017).

This city set its exact implementation date. As shown in Figure 3, in the staggered roll-out, from October 2013, all county-level hospitals implemented a new price menu, removing drug mark-up and raising prices of medical services. From December 2016, all the city-level hospitals implemented a new price menu. During the observation period from June 2014 to June 2018, only the five city-level hospitals were subject to the policy change (treated hospitals), while the other 23 hospitals had already implemented the policy (control hospitals). In this period, the prices were fixed by the government except for this policy change. The physicians' commission percentage remained unchanged as well.

The price changes were designed by the higher-level provincial government to ensure that, on average across the province, the price increases for medical procedures would offset approximately 80 percent of the revenue loss from the removal of drug mark-ups, assuming no strategic responses.Since it was a provincial-level price menu change, the sample region took the price changes as given: surgery prices in the treated hospitals went up by 30 percent. The prices of daily hospital care, such as bed, routine check, and general nursing, went up by 50 percent.

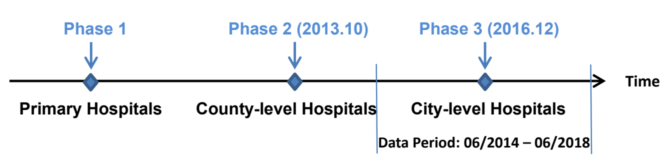


Figure 3: Policy Roll-out in Studied City

Aggregation over individual procedure price changes into treatment level suggests that surgical treatment became relatively more expensive than non-surgical treatment in treated hospitals.[20] The increase in charge differential between surgical and non-surgical treatments indicates a more intense conflict of interest between the physicians and patients: surgical treatment is more attractive to the physicians because of a higher commission amount, but less so to the patients due to a higher out-of-pocket price.

Patients are aware of such policy change and can usually have good estimates of the treatment costs. Local governments typically announce price menu changes one month in advance through official websites, social media, and TV or radio news. Moreover, physicians in China can typically provide a rough estimate of a patient's medical expenditure for various treatments, as pricing is standardized by the government and consistent across different insurance types. Physicians are usually well-versed in these price schedules. Besides, price listings are easily accessible through websites and rolling screens in hospital lobbies for patients and their families.

3.3 Data

I use de-identified claims data for hospital inpatient services from the study region. The data come from the Healthcare Security Administration of that city.

The hospital inpatient records cover all 28 hospitals in the sample region of that city and encompass ten quarters prior to the policy change and six quarters afterward. Each record is time-patient-hospital specific. It includes patient ID, hospital ID, admission and discharge dates, diagnosis name, ICD-10 code[21], total costs by procedure category (i.e., medicine, traditional Chinese medicine, surgery, materials, lab, facility, nursing, etc.), out-of-pocket price to patient, insurance payment totals, patient insurance type, patient coinsurance rate, age, gender, whether a patient is below the poverty line, whether a patient receives subsidies, and whether a patient has comorbidity.

I focus on the data for a particular diagnosis, cervical spondylosis, commonly known as neck arthritis.[22] For this diagnosis, surgical treatment refers to an episode of care including surgery, along with the surgery's related tests, medications, and hospital stay. Non-surgical treatment usually consists of physical therapy, epidural steroid injections, medications, related tests, and hospital stay. I therefore define a treatment mode indicator which is equal to 1 if surgery is among the procedures a patient undergoes. Note that non-surgical cases are sometimes admit-

[20] See Appendix D for detailed explanation.

[21] ICD-10 includes alphanumeric codes published by the World Health Organization to represent diagnoses.

[22] Neck arthritis refers to the degeneration of the discs and joints in the neck or cervical spine.

ted to the hospital, typically involving severe pain or mobility issues requiring close monitoring or combined interventions. Unlike in the U.S., where non-surgical treatments are almost exclusively outpatient, patients in China often remain hospitalized for more than a week. This difference is driven by fee-for-service payment models, hospital practices, and patient preferences for comprehensive inpatient care.

Surgical and non-surgical treatments are often substitutable. Inpatient cases frequently involve patients who have repeatedly pursued non-surgical options in the outpatient setting without achieving remission. In the data, surgeries typically occur within the first two days of admission, with no preceding non-surgical treatments, such as epidural steroid injections. The treatment choices are also discretionary. The evidence regarding the effectiveness of the non-surgical and surgical treatments is mixed from the limited medical literature about neck arthritis. There are no clear guidelines for selecting patients who may benefit from the surgery (Rowland, 1992). In my data, the expected 30-day readmission rate is the same for surgical and non-surgical treatment, conditional on patient characteristics.[23] Besides, neck arthritis is most frequently observed among diagnoses for which treatment decisions are considered discretionary, including kidney stone and prolapsed lumbar intervertebral disc.

I only include the sample for which the primary diagnosis is neck arthritis. The diagnosis accuracy rate of neck arthritis is over 95 percent with the help of CT scan or MRI.[24] I also check whether there were any malpractice lawsuits for missing or incorrect diagnoses and treatments regarding this condition on the China Judgment Online website, which contains the universe of malpractice lawsuits.[25] Among 65 malpractice cases during the four years of observation, none was about neck arthritis. The analysis sample also omits visits with a prior stay for neck arthritis within the past 120 days or visits with a length of stay exceeding 30 days.

To understand the change in the average charge differential between the two treatment modalities, I view each treatment as a bundle of procedures and borrow the idea of the Laspeyres price index by calculating the new total charges for each treatment using the original bundles and price changes. The charge differential increased by 17 percent after the policy (Appendix D provides details for calculating the change in charge differential).

[23]Additionally, my consultation with three orthopedic professionals and five Ph.D. students specializing in orthopedics in China reassures the substitutability; for few patients with severe pain and damage, surgery outperforms non-surgical treatment. Otherwise, the value of surgery depends on a physician's evaluation of the degree to which the spinal nerve is being pinched by a herniated disk, and the extent to which the spinal cord is being compressed. One concern may be that without the medical consensus on treatment selections, physicians' biased beliefs about the value of surgery for a patient may play a role in the recommendation. To mitigate this concern, I add hospital fixed effects in my empirical model to control the hospitals' potential specialty or skill differences, since physicians in the same department of a hospital work closely as a group.

[24]https://www.ncbi.nlm.nih.gov/books/NBK551557/

[25]https://wenshu.court.gov.cn/

To measure patients' general health conditions, I create a comorbidity dummy which equals 1 if there are one or more documented secondary diagnoses.[26] To measure health outcomes, I define a 30-day neck arthritis-specific readmission dummy. Given that a single patient's inpatient records can be linked over time, I set the dummy equal to 1 if I observe a patient seeking inpatient services with a primary diagnosis of neck arthritis within the following 30 days across hospitals. For the purpose of a robustness check, I also define a 90-day neck arthritis-specific readmission dummy and a 30-day all-cause readmission dummy.

Table 1: Summary Statistics for Neck Arthritis

Variable	**Mean**	**SD**	**25 Perc**	**50 Perc**	**75 Perc**
Demographics					
Age	54.65	13.25	46	54	64
Gender (male = 1)	0.38	0.48	-	-	-
Insurance type (urban = 1)	0.52	0.50	-	-	-
Coinsurance rate	0.20	0.10	0.15	0.17	0.25
Subsidy (yes = 1)	0.22	0.42	-	-	-
Comorbidity (yes = 1)	0.34	0.47	-	-	-
Treatment Outcomes					
Treatment Mode (surgery = 1)	0.08	0.28	-	-	-
Treatment Charge ($ in thousands)	1.51	1.37	0.87	1.27	1.80
-Surgical	3.03	3.41	1.70	2.03	2.73
-Non-surgical	1.37	0.87	0.84	1.21	1.63
Readmission	0.01	0.11	-	-	-
Hospitals					
Treated Hospital (yes=1)	0.28	0.45	-	-	-
Observations			10,596		

Table 1 contains summary statistics for a total of 10,596 visits. Regarding patient demographics, the average patient age is 55, and there are more female patients than males. About half of the patients are more insured. The average coinsurance rate is 20 percent. 22 percent of the patients are eligible for the government subsidy on the out-of-pocket cost. One or more comorbidity conditions are present among 34 percent of the patients. In terms of care utilization, 8 percent of the visits are surgical. The average treatment cost for neck arthritis is $1,514.[27] Surgical treatment costs over twice as much as non-surgical; with very few exceptions, the former is consistently more expensive than the latter. The 30-day readmission rate is 1 percent. 28 percent of the visits happen in the treated hospitals.

[26] I choose not to generate a comorbidity index because hospitals seem to differ substantially in terms of the way they have documented secondary diagnoses. For example, there is at most one co-existing condition for a patient in hospital A, while in hospital B, I observe four co-existing conditions.

[27] I use purchasing power parities (PPP) to transform CNY to USD. The PPP metric indicates around 3.5 CNY per USD starting from 2010.

Table 2 presents the statistics separately for the period before and after the policy change, and separately for the more insured and less insured. Overall, the demographics and market share for treated hospitals are stable across time, as measured by the normalized differences.[28] The balance check based on insurance type implies that the observed patient characteristics are not associated with insurance type.

Table 2: Balance Table for Patients with Neck Arthritis

	Before		**After**		**Normalized**	**Less Insured**		**More Insured**		**Normalized**
Variable	**Mean**	**SD**	**Mean**	**SD**	**Diff**	**Mean**	**SD**	**Mean**	**SD**	**Diff**
Age	54.03	13.25	55.26	13.23	0.09	56.06	13.31	53.37	13.06	0.20
Gender (male = 1)	0.37	0.48	0.38	0.49	0.03	0.37	0.48	0.38	0.49	0.03
Subsidy (yes = 1)	0.24	0.43	0.21	0.41	0.06	0.27	0.44	0.18	0.39	0.20
Comorbidity (yes = 1)	0.35	0.48	0.34	0.47	0.02	0.35	0.48	0.34	0.47	0.02
Treated Hospital (yes=1)	0.27	0.44	0.28	0.45	0.03	0.23	0.42	0.31	0.46	0.18
Observations	5,210		5,386			5,065		5,531		

Appendix C displays the characteristics by treated and control hospitals before the policy change. The pre-policy surgery rate is higher in the control hospitals, probably because they had implemented the price menu change before the data observation period. The fact that the treated and control hospitals differ in levels of treatment intensity and patient demographics requires more justifications of the parallel trend assumption. I therefore conduct various sensitivity analyses to support the internal validity of the Difference-in-Differences (DID) research design.

4 Direct Evidence of Financial Incentives

This section serves two purposes. First, I test the model implications. Second, I show that treatment decisions change in response to separate sources of variation in physicians' and patients' financial incentives, which serve as the key to precisely estimating the structural parameters that govern the magnitude of each financial incentive.

4.1 The Effect of Price Menu Change

I investigate whether the price menu change led to changes in the probability of surgery, treatment charge, and health outcome by using an event study framework as follows:

$$Y_{iht} = \alpha_h + \eta_t + I_i + \sum_{k=-9}^{-1} \beta_k \cdot D_h \cdot Q_k + \sum_{k=1}^{6} \beta_k \cdot D_h \cdot Q_k + \boldsymbol{\lambda} \cdot \mathbf{X_{it}} + \epsilon_{iht}, \qquad (12)$$

[28] As a rule of thumb, normalized differences over 0.25 are considered economically significant (Imbens, 2015; Imbens and Rubin, 2015).

where Y_{iht} is an outcome measure for inpatient case i in hospital h during quarter-year t. α_h and η_t denote hospital fixed effects and quarter-year fixed effects respectively. I_i is the insurance fixed effect. D_h is a dummy which equals one if hospital h is a treated hospital. Q_k is an indicator variable for the k^{th} quarter-year relative to the time of the policy. $\mathbf{X_{it}}$ is a vector of patient demographics, including patient age, gender, whether a patient receives subsidies, and whether a patient presents comorbidity. ϵ_{ith} is the error term. I omit the quarter immediately preceding the policy change so that the outcomes are relative to that quarter. The analysis extends from 10 quarters before to 6 quarters following the policy.

In addition to equation (12), I also estimate an equation in which the six post-policy quarter indicators are replaced by a single post-policy indicator $Post_t$ to capture the average treatment effect for the treated. To examine care utilization outcomes, I choose Y_{iht} to be the indicator of surgical treatment and treatment charges respectively. For health outcomes, I use the 30-day readmission dummy.

Figure 4 shows the dynamic effects ($D_h \times Q_k$) by plotting the estimated β_k coefficients with 95 percent confidence intervals for surgery probability (panel A), treatment charge (panel B), and readmission probability (panel C). It also displays the average effect ($D_h \times Post_t$).

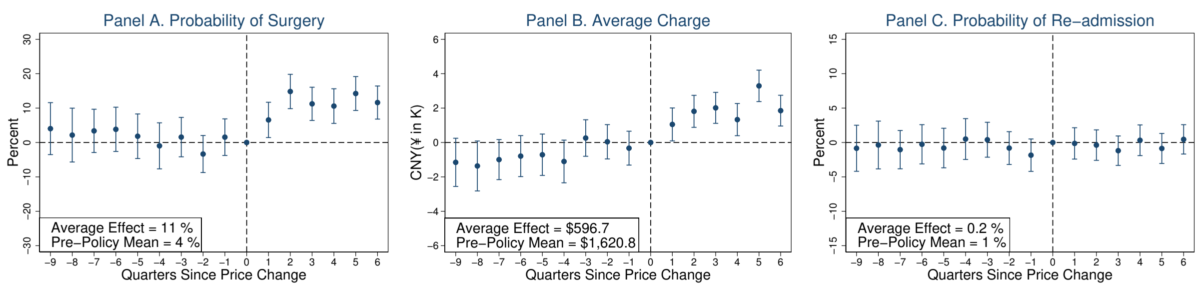


Figure 4: Effect of Price Change on Health Care Utilization and Outcome: Event Study

Note: *The figure plots regression coefficients for the policy change and their 95 percent confidence intervals from equation (12). Effects are normalized to the end of the quarter just before the policy. Standard errors are clustered at the hospital level.*

For all the measures, none of the pre-policy coefficients (β_k with $k < 0$) is statistically significant, which suggests parallel trends between the treatment and control hospitals before the policy. To the extent that the average charge shows a somewhat (non-significant) upward pre-trend, a joint-F test of the compound null, in which all the pre-policy coefficients are jointly zero, fails to reject the null hypothesis. Moreover, I have augmented equation (12) with a group-specific linear time trend. Results reported in Appendix E.2 imply a non-significant pre-trend.

Following the policy's implementation, there was a significant rise in the surgery probability, and it persisted for at least one year and a half. During the first six quarters after the policy, the surgery probability rose by 11 percentage points, or 275 percent. Consequently, panel B shows a sharp and persistent increase in the treatment charge per inpatient stay by 37 percent. A natural question raised by these facts is whether changes in care utilization influence patient health. I find no evidence of a change in health outcomes measured by the 30-day readmission rate. Since only 1 percent of patients in this sample were readmitted within 30 days, the insignificant change in the 30-day readmission rate could arise due to a lack of variation in readmission. Therefore, in Appendix E.1, I examine the policy impact on health outcome based on the 90-day neck arthritis-specific readmission and 30-day all-cause readmission dummies, and the results are robust to these alternative definitions of readmission. Of course, these results on readmission probabilities do not speak to other health effects that would not necessarily be revealed through a change in readmissions, such as patients' pain level or the duration of their pain. However, at minimum, the findings suggest that the marginal patients were not better off under surgical treatment, at least based on readmission statistics.

To gain a better understanding of the patients who switched to surgery due to the policy, I compare the patients with surgical treatment pre-policy to those in the post-policy period. As shown in Table 3, the surgery patients in the post-policy period were, on average, 1.5 years older, with 7 percentage points higher probability to receive a subsidy, and 6 percentage points more likely to present comorbidity, as compared to the surgery patients in the pre-policy period. There is no significant gender difference between the two groups.

Table 3: Surgery Patients: Pre- and Post-Policy

	Age	Gender (male=1)	Subsidy (yes=1)	Comorbidity (yes=1)	Number of Surgeries
Pre Period	53.05	0.34	0.07	0.29	309
	(10.69)	(0.48)	(0.25)	(0.46)	
Post Period	54.54	0.37	0.14	0.35	570
	(11.21)	(0.48)	(0.35)	(0.48)	
Post - Pre	1.49	0.02	0.07	0.06	
	(0.77)	(0.03)	(0.02)	(0.03)	

How significant are the policy effects? I transform the results into the elasticity of surgery probability w.r.t. physicians' incentives. To do so, I consider and calculate the change in charge differential as the measure of physicians' incentives. Appendix D shows that the charge differential increased by 17 percent after the policy. Scaling the 275 percent increase in surgery probability by this number, the elasticity of surgery probability is 16. The rise in the surgery rate is substantial. One possible explanation is that transparent protocols and information regarding the necessity of surgery are not easily accessible. Additionally, even if physicians slightly lower their thresholds for recommending surgery, a large portion of patients may be misdirected due to a thin tail in the distribution of surgery benefits (i.e, ξ_i in the model).

My findings are consistent with other studies evaluating the care utilization change due to the price menu change.[29] Fang et al. (2021) evaluates the same policy in a western Chinese county using administrative data similar to mine. They find that at the admission level, across all diagnoses, hospitals almost fully offset the reductions in drug expenditure by increasing the provision of non-drug services, with no measurable improvement in health quality. Fu et al. (2018) exploits nationally representative hospital-level data and reaches a similar conclusion.

Threats to Identification. The key to identifying the impact of physicians' financial incentives is to rule out other channels that could change surgery rates after the policy. I am primarily concerned about patients sorting into different hospitals due to the policy based on observed and unobserved preferences. For example, price-sensitive patients who are inclined to surgery disproportionally selecting hospitals with lower surgery prices could partially explain the observed change in surgery rate after the policy.

To alleviate this concern, I test for compositional change in the treated hospitals, with the idea being that patient selection based on the observables should lead to changes in the distribution of covariates. Appendix E.3 suggests no significant changes in patients' observables in the treated hospitals, except the rate of comorbidities which changes by 4 percentage points, a non-substantial magnitude given its mean of 35 percent before the policy.

To mitigate the concern of selection based on unobservables, I report four pieces of evidence. First, the surgery rates in control hospitals remained stable after the policy. If there were patient selection based on price sensitivity, there should have been a higher surgery rate after the policy change in the control hospitals where the surgery became relatively cheaper. Second, the policy change did not change the total patient volume in the treated hospitals, and the market share of treated hospitals was 27 percent before and 28 percent afterward. However, patient selection should imply a large disparity in the market share across time. Third, the policy did not change the market share of the treated hospitals for age-related cataracts either — a disease for which surgery is the only effective option and therefore there should have been many more patients switching under the selection model.[30] Lastly, for all diagnoses, patients who always went to treatment (control) hospitals rarely switched to the other hospital group after the policy. This

[29]The key of such a price menu change is to remove the 15 percent drug mark-up. The extent of the increase in other medical prices may be slightly different based on local government policy.

[30]For cataracts, when wearing different eyeglasses or using stronger lighting are no longer helpful, surgery is the only effective treatment available, according to https://www.hopkinsmedicine.org/health/conditions-and-diseases/cataractstreatment

observation is inconsistent with the switching behavior in the selection model. Details of the analysis are in Appendix E.3.

Overall, the observations reinforce the notion that patients have limited price sensitivity when choosing a hospital, compared to their sensitivity of the travel time to a given hospital (Gowrisankaran et al., 2015).

In addition to patient sorting, if hospitals admit more patients who would have stayed in outpatient care, the surgery rate could change. However, the data show no significant change in inpatient volume or market share in the treated hospitals and thus suggest minimal substitution for outpatient care (see Appendix E.3). Such results also hold by insurance types, suggesting there was no differential selection into admission.

To the extent that the increase in surgery rate reflects not only a direct response to physicians' financial incentives, but also the physicians' advanced specialization in surgery due to its complementarity with enhanced hospital capacity or technology after the policy change, one would expect higher and higher surgery rates post-policy because it takes time to build up the capacity and skills. The dynamic effects, however, are relatively flat.

4.2 Heterogeneous Effects by Insurance Type

The rise in surgery probability due to the price menu change suggests that physicians respond to their financial incentives. Here I exploit the heterogeneous responses across the more and less insured to identify patients' financial incentives. I interact the insurance dummy with the treatment effect:

$$\mathrm{Y}_{iht} = \alpha_h + \eta_t + I_i + \rho_1 \cdot I_i \cdot D_h \cdot Post_t + \rho_2 \cdot D_h \cdot Post_t + \rho_3 \cdot I_i \cdot Post_t + \rho_4 \cdot I_i \cdot D_h + \boldsymbol{\lambda} \cdot \mathbf{X_{it}} + \epsilon_{ith}. \tag{13}$$

The parameters of interest are ρ_1 and ρ_2. ρ_2 captures the average treatment effect of the policy for the less insured. ρ_1 suggests the extent to which the effects of the policy on the more insured are different from those on the less insured.

Table 4 shows the average effects for each insurance type and their differences. As indicated by the first column, the probability of surgery rose by 8 percentage points, or 200 percent, for the less insured, and 12 percentage points, or 240 percent, for the more insured due to the policy. The rise in surgery probability for the more insured is 1.5 times greater over the analysis periods. Column 2 indicates that the treatment charge per visit rose by 33 percent for the less insured and 39 percent for the more insured. Despite the greater rise in surgery probability and costs, I find no significant health improvement for either the more insured or less insured, as measured by the readmission rate in column 3.

Table 4: Heterogeneous Effects of Price Menu Change: Utilization and Health Outcome

	Surgery Probability (1)	Treatment Charge (thous. $) (2)	Readmission Probability (3)
Average Effect for Less Insured:			
$(\hat{\rho}_2)$	0.08	0.52	0.002
	(0.02)	(88.89)	(0.002)
Average Effect for More Insured:			
$(\hat{\rho}_1 + \hat{\rho}_2)$	0.12	0.64	0.002
	(0.01)	(77.60)	(0.006)
Average Effect Difference:			
$(\hat{\rho}_1)$	0.04	0.12	0.000
	(0.02)	(62.47)	(0.008)
Pre-Implementation Mean:			
Less Insured (I=0)	0.04	1.59	0.006
More Insured (I=1)	0.05	1.67	0.016

Note: *The table shows the extent to which the effects of price menu change are different between the two insurance groups, implied by equation (13). The number of observations is 10,596. Standard errors are in parentheses and clustered at the hospital level.*

Figure 5 shows the dynamic effects of the policy change by insurance type, by augmenting equation (12) with three-way interactions among the insurance dummy, the treated hospital indicator, and the time dummies and all the two-way interactions with these variables. Consistent with the average effect, it indicates a more pronounced response to the policy change for the more insured, although the differential effects are not significant in some post-policy periods due to the sample size.

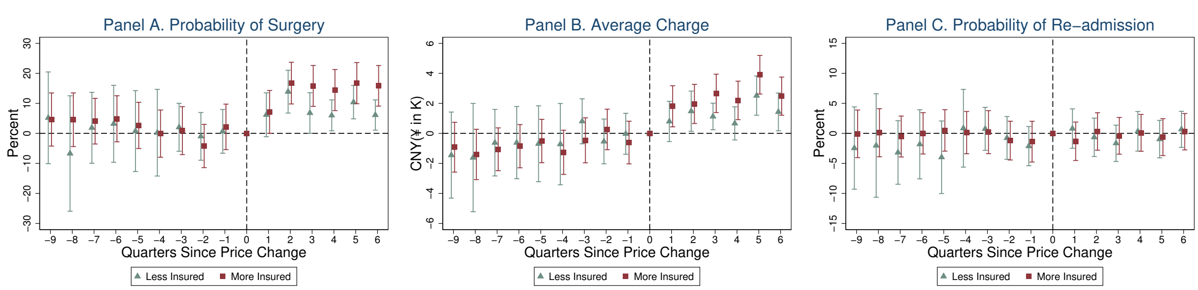


Figure 5: Heterogeneous Effect of Price Change on Health Care Utilization and Outcome

Note: *The figure plots coefficients for the policy change and their 95 percent confidence intervals for the less insured and more insured, respectively. Effects are normalized to the end of the quarter just before the policy. Standard errors are clustered at the hospital level.*

One concern is that if unobserved patient characteristics that affect surgery propensity were correlated with insurance types, the estimated difference by insurance types would pick up the effects of such factors and could not represent the causal impact of the coinsurance rates. To address this concern, I show that insurance type is not associated with a large set of variables regarding patients' health conditions and preferences. First, the treatment effect estimates for each type remain highly similar when excluding patient characteristics in equation (13); this stability of coefficients, together with the balance check in Table 2, implies uncorrelation between the insurance types and the observable characteristics. The analysis on observables is informative on the unobservables, as long as the two sets of variables are correlated. Second, I leverage external data to show that insurance type is not associated with other variables representing patients' underlying health conditions, their willingness to consume medical services or follow physicians' advice, or their risk attitudes. Details are in Appendix E.4.

The nature of the utilization response to the policy change critically depends on the interplay of physician and patient incentives. For neck arthritis, a positive and high elasticity of surgery probability, despite patients paying more out-of-pocket for surgery after the policy, suggests a more salient role of physicians' incentives in treatment choices. In inpatient settings, patient moral hazard tends to be small (Aron-Dine et al., 2013), and thus it is reasonable to expect the dominance of physician incentives in elective procedures. Indeed, for other discretionary inpatient treatment of diagnoses such as kidney stones and varicose veins, Appendix E.5 indicates a magnitude of elasticity of the probability of surgical treatment comparable to that for neck arthritis. In contrast, as a placebo test, Appendix E.5 shows that diagnoses with less discretionary treatment — for instance, kidney failure, cancer, and hip fracture, along with emergency room visits — have much lower (positive) elasticity of surgical treatment probability. However, in settings where patient moral hazard can be a major driving force, such as office visits to specialists, the policy change may imply a decrease in service volume.

To summarize, the policy change had heterogeneous effects by insurance types, consistent with model Property 1. The magnitude of the difference in responses by insurance types depends on both the patients' price elasticity and the degree of physician misdirection based on cost-sharing. However, decomposing these two forces is beyond the scope of the DID analysis and requires estimating the structural model.

4.3 Hump-shape Relationship between Surgery Probability and Conflict of Interest

The fact that the overall surgery rate rose due to the policy change implies that some physicians are unconstrained, being able to direct more patients into surgery when their relative remuneration went up even though patients faced a higher relative out-of-pocket price. This is consistent with model Property 2 when the obedience constraint does not bind.

In order to provide further suggestive evidence of Property 2, which states that the physician's cutoff (surgery probability) is decreasing (increasing) and then increasing (decreasing) with the charge differential ΔT_i, I examine surgery probability against ΔT_i for a patient with mean values for demographics conditional on particular hospital quality.

I can not directly observe ΔT_i, the potential outcome for patient i. Instead, I observe either T_i^s or T_i^d for patient i, depending on the patient's treatment decision. To be accurate, I add subscripts i, h, and q to indicate that the charge can vary across individuals, hospitals, and time. I measure the charge differential ΔT_{ihq} through the following regression

$$\begin{aligned} T_{ihq} = &\boldsymbol{\delta_1} \cdot \boldsymbol{\tau_h} + \boldsymbol{\delta_2} \cdot \boldsymbol{\psi_q} + \boldsymbol{\delta_3} \cdot \mathbf{X_{1i}} \\ &+ \boldsymbol{\delta_4} \cdot \boldsymbol{\tau_h} \cdot \mathbb{1}(\boldsymbol{s_i} = \mathbf{1}) + \boldsymbol{\delta_5} \cdot \boldsymbol{\psi_q} \cdot \mathbb{1}(\boldsymbol{s_i} = \mathbf{1}) + \boldsymbol{\delta_6} \cdot \mathbf{X_{1i}} \cdot \mathbb{1}(\boldsymbol{s_i} = \mathbf{1}) + \nu_{ihq}, \end{aligned} \tag{14}$$

where $\boldsymbol{\tau_h}$ is a vector of hospital dummies, $\boldsymbol{\psi_q}$ is a vector of quarter-year dummies, and $\boldsymbol{X_{1i}}$ is a vector of observed patient characteristics, including $\boldsymbol{X_i}$ and coinsurance rate. I also include their interaction terms with the dummy of taking surgery $\mathbb{1}(\boldsymbol{s_i} = \mathbf{1})$. The $\boldsymbol{\delta}$s are coefficients to be estimated. Eventually, I infer the charge differential as $\widehat{\Delta T_{ihq}} = \hat{\boldsymbol{\delta}}_\mathbf{4} \cdot \boldsymbol{\tau_h} + \hat{\boldsymbol{\delta}}_\mathbf{5} \cdot \boldsymbol{\psi_q} + \hat{\boldsymbol{\delta}}_\mathbf{6} \cdot \mathbf{X_{1i}}$. Appendix Figure F.8 plots the distribution of $\widehat{\Delta T_{ihq}}$.

Replacing ΔT_{ihq} with $\widehat{\Delta T_{ihq}}$, I run the following probit regression to get the relationship between the surgery probability and the charge differential:

$$\begin{aligned} Pr(s_i = 1) = \Phi\big(&\alpha_h + \eta_t + \boldsymbol{\pi}_1 \cdot \mathbf{X_i} + \pi_2 \cdot y_i \cdot \widehat{\Delta T_{ihq}} + \pi_3 \cdot \widehat{\Delta T_{ihq}} \\ &+ \pi_4 \cdot \widehat{\Delta T_{ihq}}^2 + \pi_5 \cdot \widehat{\Delta T_{ihq}}^3\big), \end{aligned} \tag{15}$$

where α_h is hospital fixed effect, η_t is time fixed effect, $\mathbf{X}_i$ is a vector of observed demographics, y_i is patient coinsurance rate, and $\widehat{\Delta T_{ihq}}$ represents the conflict of interest. Higher-order terms of $\widehat{\Delta T_{ihq}}$ are included to explore the potential non-linear relationship.

Using $\widehat{\Delta T_{ihq}}$ may cause endogeneity in the probit regression due to the potential measurement error in $\widehat{\Delta T_{ihq}}$ from regression (14). The measurement error may come from sample selection bias: surgery is not randomly assigned; there might be other omitted demographics, such as patient risk aversion and education, affecting surgery decisions. As a result, the omitted parts correlate with $\widehat{\Delta T_{ihq}}$ and the probability of surgery. In other words, the cross-sectional variations in the charge differential, due to unobserved heterogeneity in individuals, are the sources of endogeneity.

To instrument for $\Delta\widehat{T_{ihq}}$, I use the DID effect, i.e., $D_h \cdot Post_t$. This interaction term represents the exogenous policy shock that affects the charge differential. I also interact the DID effect with coinsurance rate, age, gender, comorbidity, and subsidy status to instrument for the other regressors that are functions of $\Delta\widehat{T_{ihq}}$. My implementation involves a control function approach that jointly estimates the parameters in equation (15) and the parameters that show the relationships between the endogenous variables and instruments. Appendix F.1 discusses assumptions and point estimates with and without instrumenting.

I fit the surgery probability along with $\Delta\widehat{T}_i$ at the averaged values of covariates. As shown in Figure 6, the hump shape implies that when the conflict of interest is mild, the probability of surgery rises as the physicians can get paid more from surgery, consistent with the policy change's overall effect. However, when the conflict of interest reaches a threshold, the probability decreases with the charge differential.

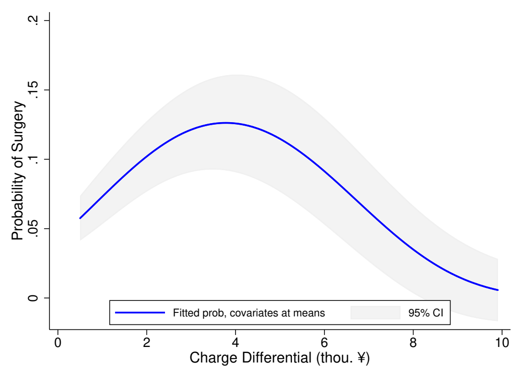


Figure 6: Predicted Surgery Probability and Conflict of Interest

Note: *Conflict of interest is measured by the charge differential. The fitted relationship is conditional on patient demographics, time trends, and hospital quality. Confidence intervals are based on bootstrap, accounting for the measurement of* ΔT_{ihq} *in Equation* (14) *and the two stages in the main regression* (15).

It might be concerning that the charge differential does not take into account the underlying costs of performing the two treatments, such as physicians' time and malpractice risk, and thus fails to accurately represent physicians' financial incentives. Appendix F.2 assesses these costs with supplemental evidence and concludes that the cost differences are limited in this setting. To the extent that the cost to physicians of performing surgical treatment is slightly greater than the cost of providing non-surgical treatment, physicians' sensitivity to financial incentives would be underestimated if such cost difference is ignored.

This hump-shaped relationship cannot be attributed to capacity constraints, as only 0.4 percent of the days in treated hospitals have a daily bed occupancy rate exceeding 90 percent. Furthermore, the differences in length of stay and physicians' actual time use between the two treatment options are minimal (Appendix F.2). Although surgical treatment results in slightly longer bed occupancy compared to non-surgical treatment (14 days vs. 12 days), I confirm in Appendix F.3 that the results remain robust to alternative specifications of the charge differential per day.

Additional Evidence Supporting the Bayesian Persuasion Model. Although no information intervention occurred during the sample period, an ideal test of the model would involve examining the effects of such an intervention to validate the role of asymmetric information. I provide three additional pieces of evidence in support of the model.

First, according to Property 2, the model also implies a "U" shape relationship between the expected value of ξ_i and the charge differential conditional on a surgery patient. Appendix F.4 uses the readmission indicator as the proxy for the unobserved benefit ξ_i and demonstrates this relationship. However, the estimates are not precise due to the lack of variation in the readmission.

Second, the Bayesian persuasion model predicts that in equilibrium patients always follow physicians' recommendations. Thus, the absence of patients rejecting recommendations provides evidence supporting the model. For neck arthritis, no patient was admitted to one hospital for a short period with little treatment and subsequently admitted to another hospital for the same condition within 10 days of the initial discharge. These inpatients likely sought outpatient care multiple times without achieving remission. By the time they require inpatient care, they are less likely to shop around for physicians or hospitals. My survey of 11 orthopedic physicians and 58 patients in China also suggests that, over 95 percent of the time, patients follow physicians' advice.

Finally, a quote from a healthcare professional reflects the model that physicians are aware of patient skepticism and take it into account when making recommendations: "There can be legitimate reasons to feel skepticism and mistrust... When addressing medical skepticism in a healthcare setting, efficient communication techniques can improve the patient relationship, making patients more willing to consider new ideas and even change their minds."[31]

5 Estimation of Empirical Persuasion Model

I use MLE to estimate a parameterized version of the model in Section 2 to study the interplay of information asymmetry and financial incentives.

[31] https://nursinglicensemap.com/blog/how-to-break-through-medical-skepticism/

5.1 Parameterization

Recall that patient i's relative utility from surgery is $U_i^s = \xi_i - \kappa_i$, where κ_i represents the common knowledge of the relative value of surgery for i,

$$\begin{aligned}
\kappa_i &= k(y_i \cdot \Delta T_{ihq}, \mathbf{Z}_i) \\
&= \underbrace{\alpha_1 \cdot y_i \cdot \Delta T_{ihq} + \alpha_2 \cdot (y_i \cdot \Delta T_{ihq})^2}_{\text{effect of out-of-pocket price}} + \underbrace{\tau_h}_{\text{hospital's quality}} + \underbrace{\psi_q}_{\text{time dummies}} \\
&\quad + \underbrace{\gamma_1 \cdot age_i + \gamma_2 \cdot gender_i + \gamma_3 \cdot subsidized_i + \gamma_4 \cdot comor_i + \gamma_5 \cdot \omega_i}_{\text{value affected by demographics}} \\
&= \alpha_1 \cdot y_i \cdot \Delta T_{ihq} + \alpha_2 \cdot (y_i \cdot \Delta T_{ihq})^2 + \tau_h + \psi_q + \boldsymbol{\gamma} \cdot \mathbf{X}_i + \gamma_5 \cdot \omega_i,
\end{aligned} \tag{16}$$

where $\mathbf{X}_i = [age_i, gender_i, subsidized_i, comor_i]$, including all patient demographics observed by researchers, $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \gamma_3, \gamma_4]$, and ω_i represents patient demographics observed by both the patient and physician, but not by researchers. κ_i is specified as a quadratic function of the out-of-pocket price to capture non-linear marginal disutility of costs.

A physician's financial incentives, F_i, is assumed to be a quadratic function of charge differential to capture potential income effects,

$$F_i = \beta_1 \cdot \Delta T_{ihq} + \beta_2 \cdot (\Delta T_{ihq})^2. \tag{17}$$

Ideally, the distribution of ξ_i should be linked to an observed patient health outcome. Appendix G provides an extension of the model, where ξ_i is the latent health benefit whose distribution can be traced out using the readmission indicator. However, due to the fact that plentiful health information would not show up in the readmission indicator and little variation in readmission exists, the estimated distribution, along with other model estimates, is imprecise. Alternatively, I assume ξ_i to be standard normal with CDF $\Phi(\cdot)$ and PDF $\phi(\cdot)$. The normality assumption is for tractability and is quite robust to distribution misspecifications in estimating mean parameters (Gourieroux et al., 1984). Assuming unit variance serves as a scale normalization for model identification. However, the distributional choice for ξ_i matters for welfare analysis, as the patient expected utility depends on the full distribution of ξ_i. With this caveat in mind, Appendix A.2 shows that the optimal cutoff under normality becomes

$$\begin{aligned}
c_i^*(\Delta T_{ihq}, y_i, \tau_h, \psi_q, \mathbf{X_i}, \omega_i) &= \max\{-\lambda^{-1}(\kappa_i),\ \kappa_i - F_i\} \\
&= \max\Big\{ -\lambda^{-1}\big(\alpha_1 \cdot y_i \cdot \Delta T_{ihq} + \alpha_2 \cdot (y_i \cdot \Delta T_{ihq})^2 + \tau_h + \psi_q + \boldsymbol{\gamma} \cdot \mathbf{X}_i + \gamma_5 \cdot \omega_i\big), \\
&\quad (\alpha_1 \cdot y_i - \beta_1) \cdot \Delta T_{ihq} + (\alpha_2 \cdot y_i^2 - \beta_2) \cdot {\Delta T_{ihq}}^2 + \tau_h + \psi_q + \boldsymbol{\gamma} \cdot \mathbf{X}_i + \gamma_5 \cdot \omega_i\Big\},
\end{aligned} \tag{18}$$

where $\lambda^{-1}(x)$ is the inverse function of inverse mills ratio $\lambda(x) \equiv \frac{\phi(x)}{\Phi(x)}$.

5.2 Separating Physician and Patient Financial Sensitivity

Separating physician and patient financial sensitivity typically requires observation of individual decisions of each party, that is, physicians' recommendations, and patients' decisions to follow or not. In my context, observing how a physician's recommendation changes with ΔT_{ihq} is key to identifying physicians' sensitivity to remuneration, β_1 and β_2, while observing how a patient's decision to accept or reject changes with coinsurance rate y_i is key to identifying patients' sensitivity to out-of-pocket price, α_1 and α_2. However, it is rare to obtain data on separate decisions in healthcare markets, except for the prescription drugs market (Sacks, 2018). Instead, my data, like other hospital datasets, only include treatment decisions.

I overcome the aforementioned data limitation by leveraging the Bayesian persuasion model and independent sources of variations in ΔT_{ihq} and y_i. Since the model implies a patient's obedience to a physician's recommendation in equilibrium, the treatment decisions perfectly predict physicians' recommendations.[32] As a result, data on treatment decisions, together with the independent sources of variation in ΔT_{ihq} and y_i, is sufficient to identify their financial sensitivity separately.

When the obedience constraint binds, the optimal cutoff is $c_i^* = -\lambda^{-1}\big(\alpha_1 \cdot y_i \cdot \Delta T_{ihq} + \alpha_2 \cdot (y_i \cdot \Delta T_{ihq})^2 + \tau_h + \psi_q + \boldsymbol{\gamma} \cdot \mathbf{X}_i + \gamma_5 \cdot \omega_i\big)$, where only the patients' sensitivity is at play. α_1 and α_2 are identified by the variation in treatment choices through varying y_i, conditional on the hospital, time, charge differential, and other patient demographics. One concern may be that if most patients within a hospital have the same *Hukou* registration, inadequate variation in y_i can be exploited. According to Table C.1 in the appendix, there is a sizable amount of coinsurance rate variation even after conditioning on hospital fixed effects.

When the obedience constraint does not bind, the cutoff becomes $c_i^* = (\alpha_1 \cdot y_i - \beta_1) \cdot \Delta T_{ihq} + (\alpha_2 \cdot y_i^2 - \beta_2) \cdot {\Delta T_{ihq}}^2 + \tau_h + \psi_q + \boldsymbol{\gamma} \cdot \mathbf{X}_i + \gamma_5 \cdot \omega_i$. With the variation in treatment choices through varying ΔT_{ihq}, β_1 and β_2 can be pinned down once α_1 and α_2 are estimated. Therefore, one condition to separately identify physician and patient financial sensitivity is to have both constrained and unconstrained physicians.

A potential concern in estimation is the endogeneity of ΔT_{ihq}. I address the potential hospital price-quality correlation by controlling for the hospital fixed effects. Conditional on hospital

[32] Although it is hard to conclude that all the patients follow empirically, there is suggestive evidence to support such equilibrium as a good approximation: for neck arthritis, no patient was admitted to one hospital and then appeared as a separate inpatient visit in another hospital for the same reason within 30 days of last discharge in the study region, undermining the concern of a second opinion.

fixed effects, replacing ΔT_{ihq} with the imputed $\widehat{\Delta T_{ihq}}$ could bring other endogenous variations due to the omitted patient demographics ω_i. The control function approach in the next section only exploits exogenous variation in $\widehat{\Delta T_{ihq}}$ due to the policy change.

5.3 Maximum Likelihood and a Control Function Approach

Since the model implies that a patient will always follow his physician's recommendation, the probability of surgery equals the probability of recommending surgery, which is the probability of ξ_i being no less than the optimal cutoff $c_i^*(\Delta T_{ihq}, y_i, \tau_h, \mathbf{X_i}, \omega_i)$. I can then write down the likelihood function.

To estimate, I need to replace ΔT_{ihq} with $\widehat{\Delta T_{ihq}}$ from regression (14) and omit the omitted ω_i. However, as discussed in Section 4.3, the measurement error of $\widehat{\Delta T_{ihq}}$ indicates a correlation between $\widehat{\Delta T_{ihq}}$ and the unobserved ω_i. To address the endogeneity in $\widehat{\Delta T_{ihq}}$, I apply a control function approach where I use the policy change, $D_h \cdot Post_t$, as the instrument in the first stage. Essentially, the estimation only relies on the variation in $\widehat{\Delta T_{ihq}}$ that comes from the policy change. The first-stage is:

$$\widehat{\Delta T_{ihq}} = \tau_h + \psi_q + \Pi \cdot [y_i,\ \mathbf{X_i},\ D_h \cdot Post_t]' + \nu_{ihq}. \tag{19}$$

I specify the control function as linear in $\widehat{\nu_{ihq}}$, the residual from the first-stage linear regression. The expression for the optimal cutoff becomes

$$\begin{aligned} c_i^*(\widehat{\Delta T_{ihq}}, y_i, \tau_h, \psi_q, \mathbf{X_i}, \widehat{\nu_{ihq}}) = \max\Big\{ & -\lambda^{-1}\big(\alpha_1 \cdot y_i \cdot \widehat{\Delta T_{ihq}} + \alpha_2 \cdot (y_i \cdot \widehat{\Delta T_{ihq}})^2 + \tau_h + \psi_q + \boldsymbol{\gamma} \cdot \mathbf{X}_i + \zeta_1 \cdot \widehat{\nu_{ihq}}\big), \\ & (\alpha_1 \cdot y_i - \beta_1) \cdot \widehat{\Delta T_{ihq}} + (\alpha_2 \cdot y_i^2 - \beta_2) \cdot \widehat{\Delta T_{ihq}}^2 + \tau_h + \psi_q + \boldsymbol{\gamma} \cdot \mathbf{X}_i + \zeta_2 \cdot \widehat{\nu_{ihq}}\Big\}. \end{aligned} \tag{20}$$

The standard errors are obtained through bootstrap, accounting for the measurement of ΔT_{ihq} from Equation (14) and the two stages in the main estimation.

6 Estimation Results

6.1 Estimation Results

Table 5 displays the estimates of the Bayesian persuasion model with and without the control function. That the coefficients for the residuals from the first stage are estimated significantly justifies the application of a control function approach.

In the Bayesian persuasion model with control function, the positive and significant $\hat{\alpha}_1$ and $\hat{\beta}_1$ suggest patients' disutility from out-of-pocket price and physicians' utility gain from charge differential. Positive $\hat{\alpha_2}$ indicates patients' increasing disutility from financial incentives, and negative $\hat{\beta}_2$ suggests physician income effects, while neither is statistically significant. Meanwhile, the financial sensitivity parameters jointly determine the degree of physicians' altruism. First, note that a physician typically gets 20 percent of the hospital revenue. Therefore, scaling up physician sensitivity parameters by the fraction of revenue, the physician's utility from x dollars income is $5\hat{\beta}_1 x + 25\hat{\beta}_2 x^2$. This amount of utility gain can be achieved by giving a patient z dollars such that $\hat{\alpha_1} z + \hat{\alpha_2} z^2 = 5\hat{\beta}_1 x + 25\hat{\beta}_2 x^2$. By valuing the patient's z dollars as x dollars for herself, the physician's weight on the patient is $\frac{x}{x+z}$. Note that the weight varies with the physician's financial incentive x, because as the physician gets paid more, her marginal utility of money decreases, and the weight on patient benefit increases. Overall, the weight ranges from 8 percent to 13 percent. Particularly, a physician facing the average charge differential trades off 1 dollar of her income for 8 dollars of patient out-of-pocket cost, placing a weight of 11 percent on patient benefit. This level of altruism is lower than that in prescription drug markets in the U.S. (around half in Schnell (2017)) and Japan (more than half in Iizuka (2007)), but higher than findings of magnitude close to zero, such as Iizuka (2012), and Crea et al. (2019).

Table 5: Structural Model Parameter Estimates

	Bayesian Persuasion (1)	(2)
Financial Incentives		
β_1 : Charge Differential	0.004	0.154
	(0.001)	(0.065)
β_2 : (Charge Differential)^2	-0.002	-0.001
	(0.001)	(0.001)
α_1 : OOP	0.080	0.065
	(0.028)	(0.033)
α_2 : OOP^2	0.004	0.003
	(0.003)	(0.005)
Patient Demographics		
γ_1 : Age	0.027	0.041
	(0.008)	(0.015)
γ_2 : Gender (male=1)	-0.139	-0.084
	(0.083)	(0.058)
γ_3 : Subsidy (yes=1)	0.019	0.074
	(0.009)	(0.062)
γ_4 : Comorbidity (yes=1)	0.002	0.014
	(0.001)	(0.006)
ζ_1 : Residual for $\Delta\widehat{T_{ihq}}$ in the First Argument of Max Function		0.106
		(0.052)
ζ_2 : Residual for $\Delta\widehat{T_{ihq}}$ in the Second Argument of Max Function		0.022
		(0.012)
Hospital Fixed Effects	Yes	Yes
Quarter-Year Fixed Effects	Yes	Yes
Instrumental Variables	No	Yes

Note: *Bootstrap standard errors in parentheses, applied to the measurement of ΔT_{ihq} in equation (14) and both stages of the maximum likelihood estimation.*

The estimates suggest that a 1 percent increase in patient coinsurance rate leads to a 0.03 percent decrease in treatment charge. This price elasticity of -0.03 reflects that both patients and physicians respond to the coinsurance rate, due to a combination of price sensitivity, physicians' altruism, and the obedience constraint in the model. Aron-Dine et al. (2013) translates the experimental treatment effects from the RAND Health Insurance Experiment into comparable arc price elasticities of -0.04 to -0.09, pooling inpatient and outpatient spending. My estimate of the moral hazard effect is smaller, primarily because the demand for inpatient surgeries is generally much less elastic than that for outpatient services. In addition, the contexts vary substantially and may lead to different estimates even for inpatient services. Notably, the sole moral hazard effect without physician response to the coinsurance rate is estimated to be -0.02. The difference in elasticities with and without physician response highlights the role of physician misdirection based on cost-sharing.

Consistent with the evidence from Section 4, the estimated γ_1 indicates that the value of surgery decreases with age. While women are more averse to surgery, this gender difference is insignificant. Moreover, people who receive subsidies are more reluctant to opt for surgical treatment, possibly because the subsidy is lump-sum and falls far short of the out-of-pocket expenditure. The effect of the subsidy status becomes insignificant when applying the control function approach. The presence of comorbidity significantly decreases the benefits of surgery.

6.2 Role of Information Asymmetry and Financial Incentives

I use the case with asymmetric information as the baseline and then shut down physicians' information advantage to quantify the extent of physician misdirection due to asymmetric information. Note that when a physician's financial incentives disappear ($F_i = 0$), her cutoff becomes $c_i^* = \kappa_i$, which is precisely the first best cutoff when the physician's private information about ξ_i becomes common knowledge and the patient makes his own choice. Therefore, shutting down physicians' financial incentives can reach the same treatment assignments as giving patients full information (while keeping physician financial incentives).

Table 6 presents comparisons between the baseline and counterfactual case. The analytical expressions for the surgery probability, total expenditure, and patient welfare can be found in Appendix H. When fully informing patients, equivalently removing physician financial incentives, 3.7 percent $(8.3 - 4.6)$ of all patients, or 45 percent of surgery patients, would not have chosen surgery. In other words, these patients are misdirected due to incentive misalignment and asymmetric information between the two parties. Furthermore, eliminating physician financial incentives saves this region 0.7 million dollars across 4 years, a 5 percent decrease in total expenditure. Correspondingly, expenditure savings per misdirected patient is nearly

\$2,000. With incentive alignment, the patient expected welfare improves by 82 percent because they save on expected treatment costs, and patients who were worse candidates for surgery, for example, those who were older and had comorbidities, would not have been misdirected into surgery.

Table 6: Extent of Physicians' Misdirection

	Baseline Model			Physician's Private Info Becomes Common Knowledge		
	All Sample	Less Insured	More Insured	All Sample	Less Insured	More Insured
Surgery Probability (%)	8.3	7.4	9.2	4.6	4.2	4.9
Percent Change Relative to Baseline				-45	-42	-47
Total Expenditure (Mill. $)	15.8	7.3	8.5	15.1	7.1	8.0
Percent Change Relative to Baseline				-5	-3	-6
Patient Welfare ($)	2,562.5	1,189.8	1,372.7	4,652.8	2,083.4	2,569.4
Percent Change Relative to Baseline				82	75	87

To show how this distortion depends on a patient's financial incentives, I revisit the above comparison for the more insured and the less insured separately in Table 6. It is more likely to misdirect a more-insured patient because more-insured patients are less price-sensitive: 3.2 percent $(7.4 - 4.2)$ of the less insured are misdirected, while 4.3 percent $(9.2 - 4.9)$ of the more insured are. As a result, there is a more substantial improvement in total savings and patient welfare for the more insured when physicians become the perfect agents for the patients.

The comparison by insurance type suggests that part of the observed differences in care utilization by coinsurance rate are attributed to greater misdirection based on cost-sharing. The intuition is that neither the patients nor the physicians internalize the insurer's cost. As a result, the lower the patients' cost-sharing, the more vulnerable the patients are to physicians' misdirection, and the larger the patient welfare loss due to the information asymmetry. The fact that insurance coverage exacerbates physician misdirection highlights a new offsetting effect on insurance value. This effect is in addition to the classical moral hazard channel, whereby patients themselves consume more when they are less price-sensitive. To further shed light on the separate mechanisms through moral hazard and misdirection, Section 6.4 decomposes the effect of cost-sharing into the two channels.

Another way to assess the importance of physician financial incentives is to evaluate the unintended consequences of the price menu change, whose designer ignored physicians' strategic behaviors stemming from financial incentives when projecting care utilization under new prices. Using the outcomes for treated hospitals in the post-policy period as a benchmark, I find that if the old price menu were in place for the treated hospitals in the post-policy period, surgery rates would decrease from 12.4 percent to 4.6 percent, the total expenditure would drop from 10.5 million to 8 million, and patient welfare would grow by over five times. While this policy aimed to contain medical costs, it eventually caused the opposite.

6.3 Counterfactual Policy Regarding Physicians' Remuneration

What happens to care utilization and patient welfare when a physician's compensation increases from its current level of 20 percent to 100 percent of the hospital revenue generated by the physician? This counterfactual exercise serves two purposes. First, it speaks to model Property 1, which predicts that physicians' responses to compensation changes depend on patient coinsurance. Therefore, I examine these effects by insurance type in Table 7. Second, it evaluates whether the Bayesian persuasion model provides reasonable counterfactual analysis by comparing them to "reality". This is because this counterfactual experiment reflects a physician remuneration reform actually implemented in China. In 2016, one of the largest hospitals in China reformed physicians' payment schemes in randomly selected departments (Gong et al., 2021). Previously, the physicians' bonus payment was 15-20 percent of their total revenue creation, similar to my setting. The reform substantially increased the dependence of physician compensation on hospital revenue, raising the share to 90 percent.

Table 7: Effect of Physician Payment Reform

	Baseline			Predictions for Policy Change in 2016			Real Change in 2016
	All sample	Less Insured	More Insured	All Sample	Less Insured	More Insured	All Sample
Surgery Probability (%)	8.3	7.4	9.2	10.1	8.7	11.3	
Percent Change Relative to Baseline				22.3	18.0	22.9	21
Total Expenditure (Mill. $)	15.8	7.3	8.5	16.0	7.4	8.6	
Percent Change Relative to Baseline				0.9	0.8	1.1	3
Patient Welfare ($)	2562.5	1189.8	1372.7	490.3	253.8	236.5	
Percent Change Relative to Baseline				-80.8	-78.6	-82.8	-

Note: *Model predictions are based on my data for neck arthritis. Numbers for real change are from Gong et al. (2021). Real total expenditure change is calculated based on reported changes in daily expenditure and length of stay.*

Table 7 confirms that, while this policy does not change patients' out-of-pocket price, patients' coinsurance rate indeed affects how physicians respond. When the dependence of physicians' remuneration on their hospital revenue increases from 20 percent to 100 percent, the surgery rate is predicted to rise by 18 percent for the less insured while by 23 percent for the more insured. Consequently, patient welfare decreases by a larger extent for the more insured.

Since the departments responsible for designing insurance policies and those focused on physician payment structures are typically separate, the incentives of both may change simultaneously. To demonstrate that designing physician incentives without considering the simultaneous changes in insurance policies can lead to unintended consequences, I simulate the effect of the 2016 physician payment reform under different assumptions about patient coinsurance rates. In Figure 7, if only physician payments are reformed, the surgery probability increases from the baseline of 8.3 percent to 10.1 percent. However, if the patient coinsurance rate decreases (increases) simultaneously, the simulated surgery probability rises (falls) compared to the case where only physician payments change. Notably, the surgery probability may become larger, smaller, or equal to the baseline rate depending on the direction of the coinsurance adjustment. This suggests that simultaneously designing incentives for both physicians and patients may be necessary to achieve the intended policy outcomes.

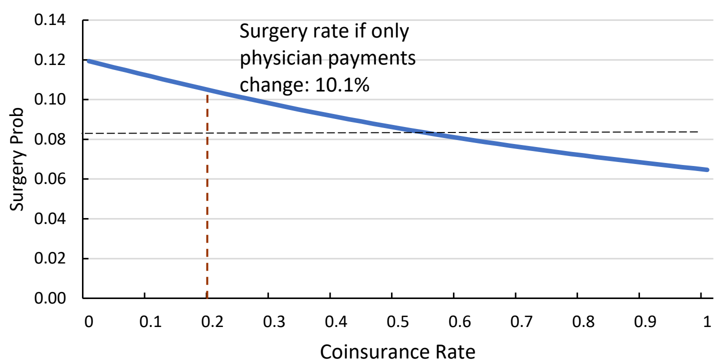


Figure 7: Effect of Physician Payment Change when Patient Coinsurance Also Changes

Note: *The figure shows that the predicted surgery probability, when physicians receive all the hospital revenue they generate, depends on whether there is a simultaneous change in patient insurance policies. In each scenario, each patient's coinsurance rate changes by 0.02.*

Finally, I compare the results to real-world data. In the last column of Table 7, Gong et al. (2021) estimates that the real 2016 physician payment reform led to a 21 percent increase in the surgery rate for randomly selected departments. Thus, the Bayesian persuasion model produces realistic counterfactual predictions.

6.4 Decomposing the Moral Hazard and Greater Misdirection Effects

This decomposition is only possible in a model that explicitly captures the patient decision-making process. To quantitatively examine the separate mechanisms of the moral hazard and greater misdirection effects, I analyze a policy that reduces patients' coinsurance rates by 50 percent. This policy is in line with the goal of most Chinese provinces. In the latest *14th Five-Year Plan for National Medical Security*, the central government also aims to steadily decrease coinsurance rates for certain services.[33] In the full-policy scenario, both effects are

[33] http://english.www.gov.cn/premier/news/202109/15/content$_{W}S6141f29ac6d0df57f98e038f.html$.

at play. To isolate the effects, in the moral hazard-only scenario, I consider the counterfactual outcomes as if the movement of the cutoff is of the distance $|\kappa_i^{'} - \kappa_i|$, where the prime symbol represents the counterfactual case. In the misdirection-only scenario, the movement of the cutoff is $|(h^{-1}(\kappa_i^{'}) - \kappa_i^{'}) - (h^{-1}(\kappa_i) - \kappa_i)|$, as in the decomposition equation (11).

Figure 8 presents the percent changes. Both traditional moral hazard and greater misdirection based on lower coinsurance rates contribute to a 9.7 percent increase in surgeries, with the misdirection effect accounting for about one-fifth of this increase. While patients experience a welfare gain of 8.3 percent due to lower out-of-pocket costs, greater misdirection reduces patient welfare by 7.3 percent, indicating that some patients are worse off due to surgeries they would not have opted for otherwise. As a result, the greater misdirection effect nearly offsets the welfare gain.

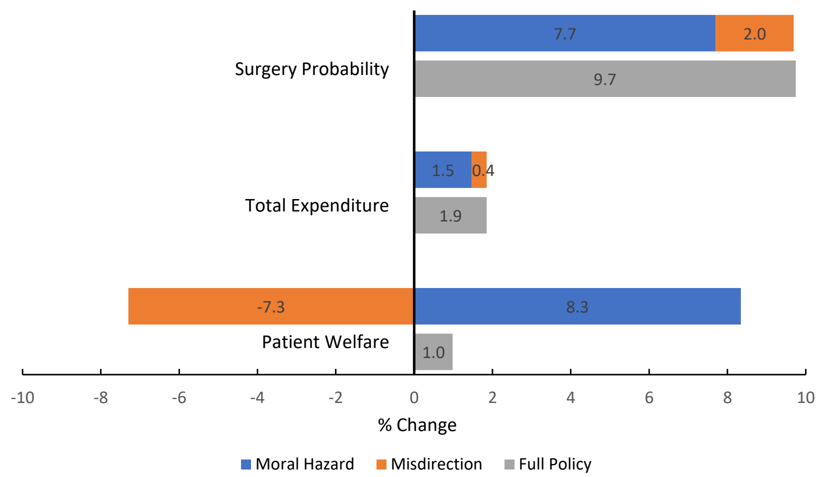


Figure 8: Decomposition When Coinsurance Rate Decreases by Half

Note: *The figure shows the percent change in outcomes under the moral hazard-only, misdirection-only, and full-policy scenarios, respectively. The baseline case is when everyone's coinsurance rate is as it is in the data. The counterfactual case is when they face half of these coinsurance rates.*

In summary, while increasing insurance generosity may initially seem to improve patient welfare by reducing out-of-pocket costs, the misdirection effect undermines the value of insurance and can even result in a net loss in patient welfare with more generous coverage.

The decomposition exercise is not meant to suggest that the Bayesian persuasion model is the ultimate framework for all incentive design policies, but rather that explicitly modeling patients' decision-making processes is crucial for certain policies. For instance, insurance policymakers should account for the new misdirection effect, which can offset the value of insurance for patients.

6.5 Discussion: The Role of Obedience Constraint

A common feature of most models of physician decision-making is the use of a physician-patient utility, which is a weighted sum of the patient's utility and the physician's utility. The Bayesian persuasion framework, on top of this utility modeling, explicitly models how skeptical patients update beliefs and make decisions, through the obedience constraint. Using model estimates, I show that the obedience constraint is empirically relevant, with 73 percent of physicians facing binding constraints when recommending surgery. Without this constraint, the probability of recommending surgery increases from 8 percent to 20 percent, assuming patients always follow recommendations.

In some settings, it may be sufficient to use a weighted-sum objective function alone, without the obedience constraint. For instance, when considering treatment decisions in emergency departments, patients have limited decision-making power. Also, if the research question focuses on insurer-physician bargaining, simplifying the physician-patient interaction may be beneficial. However, to address questions about the interaction of incentives and their interplay with information asymmetry, such as those in this paper, I demonstrate that explicitly modeling the patient's active role is necessary. In Appendix I, I discuss the implications of a model without the obedience constraint, estimate this alternative model, and compare the counterfactual analysis between the Bayesian persuasion model and the model without obedience constraint.

7 Conclusion

A physician and her patient decide jointly on medical treatment. I apply the theory of Bayesian persuasion to study the physician-patient relationship, explicitly considering the patient's decision-making process.

I exploit a setting in China, wherein treatment decisions vary with the independent sources of variation in physicians' remuneration and patients' coinsurance rates, to disentangle the role of physicians' financial incentives from their patients' in the model. The counterfactual exercise suggests that 45 percent of the surgeries are unnecessary. A larger fraction of the more insured are misdirected. This greater misdirection effect highlights a new channel that works to offset the value of insurance in addition to the traditional moral hazard.

In future work, I hope to have more detailed data on health outcomes to better understand the effects of the policy change and to implement the framework outlined in Appendix G, integrating health outcomes into the welfare analysis.

Building on the current setup, the Bayesian persuasion model can be extended to incorporate additional features, such as patients' uncertainty about physicians' financial incentives or patients' biased beliefs. While these factors present promising directions for future research with appropriate data, their estimation is beyond the scope of this paper.

References

Aron-Dine, A., L. Einav, and A. Finkelstein (2013). The rand health insurance experiment, three decades later. *Journal of Economic Perspectives 27*(1), 197–222.

Aumann, R. J., M. Maschler, and R. E. Stearns (1995). *Repeated games with incomplete information*. MIT press.

Baicker, K., S. Mullainathan, and J. Schwartzstein (2015). Behavioral hazard in health insurance. *The Quarterly Journal of Economics 130*(4), 1623–1667.

Bergemann, D. and S. Morris (2019). Information design: A unified perspective. *Journal of Economic Literature 57*(1), 44–95.

Berry, S. T. (1992). Estimation of a model of entry in the airline industry. *Econometrica: Journal of the Econometric Society*, 889–917.

Burkhardt, J.-K., A. F. Mannion, S. Marbacher, P. A. Dolp, T. F. Fekete, D. Jeszenszky, and F. Porchet (2013). A comparative effectiveness study of patient-rated and radiographic outcome after 2 types of decompression with fusion for spondylotic myelopathy: anterior cervical discectomy versus corpectomy. *Neurosurgical focus 35*(1), E4.

Cabral, M., C. Carey, and S. Miller (2021). The impact of provider payments on health care utilization of low-income individuals: Evidence from medicare and medicaid. Technical report, National Bureau of Economic Research.

Chan, K. W. and L. Zhang (1999). The hukou system and rural-urban migration in china: Processes and changes. *The China Quarterly 160*, 818–855.

Chan Jr, D. C., M. Gentzkow, and C. Yu (2019). Selection with variation in diagnostic skill: Evidence from radiologists. Technical report, National Bureau of Economic Research.

Chandra, A., D. Cutler, and Z. Song (2011). Who ordered that? the economics of treatment choices in medical care. *Handbook of health economics 2*, 397–432.

Chen, A. and D. N. Lakdawalla (2019). Healing the poor: The influence of patient socioeconomic status on physician supply responses. *Journal of health economics 64*, 43–54.

Chernew, M., T. B. Gibson, K. Yu-Isenberg, M. C. Sokol, A. B. Rosen, and A. M. Fendrick (2008). Effects of increased patient cost sharing on socioeconomic disparities in health care. *Journal of general internal medicine 23*(8), 1131–1136.

Choi, D., M. J. Vives, and S. B. Chaudhary (2016). Cervical spondylotic myelopathy: Treatment options. *Recent Advances in Spinal Surgery 117*.

Clemens, J. and J. D. Gottlieb (2014). Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review 104*(4), 1320–49.

Crea, G., M. M. Galizzi, I. Linnosmaa, and M. Miraldo (2019). Physician altruism and moral hazard:(no) evidence from finnish national prescriptions data. *Journal of health economics 65*, 153–169.

Currie, J. and W. B. MacLeod (2017). Diagnosing expertise: Human capital, decision making, and performance among physicians. *Journal of labor economics 35*(1), 1–43.

Decker, C. (2022). What's in an airbnb five-star rating? an empirical model of bayesian persuasion.

Dickstein, M. (2011). Physician vs. patient incentives in prescription drug choice.

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

Einav, L., A. Finkelstein, and N. Mahoney (2018). Provider incentives and healthcare costs: Evidence from long-term care hospitals. *Econometrica 86*(6), 2161–2219.

Eliason, P. J., P. L. Grieco, R. C. McDevitt, and J. W. Roberts (2018). Strategic patient discharge: The case of long-term care hospitals. *American Economic Review 108*(11), 3232–65.

Ellis, R. P. and T. G. McGuire (1986). Provider behavior under prospective reimbursement: Cost sharing and supply. *Journal of health economics 5*(2), 129–151.

Fang, H., X. Lei, J. Shi, and X. Yi (2021). Physician-induced demand: Evidence from china's drug price zero-markup policy. *NBER Working Paper*.

Feldman, R. and B. Dowd (1991). A new estimate of the welfare loss of excess health insurance. *The American economic review 81*(1), 297–301.

Feldstein, M. (1977). Quality change and the demand for hospital care. *Econometrica: journal of the Econometric Society*, 1681–1702.

Finkelstein, A. (2007). The aggregate effects of health insurance: Evidence from the introduction of medicare. *The quarterly journal of economics 122*(1), 1–37.

Fu, H., L. Li, and W. Yip (2018). Intended and unintended impacts of price changes for drugs and medical services: Evidence from china. *Social Science & Medicine 211*, 114–122.

Geddes, E. and M. Schnell (2023). The expansionary and contractionary supply-side effects of health insurance. Technical report, National Bureau of Economic Research.

Gong, Q., L. Li, X. Liu, and X. Zhao (2021). Financial incentives and inpatient care: Evidence from a randomized physician remuneration reform. *Working Paper*.

Gourieroux, C., A. Monfort, and A. Trognon (1984). Pseudo maximum likelihood methods: Theory. *Econometrica: journal of the Econometric Society*, 681–700.

Gowrisankaran, G., A. Nevo, and R. Town (2015). Mergers when prices are negotiated: Evidence from the hospital industry. *American Economic Review 105*(1), 172–203.

Gruber, J., J. Kim, and D. Mayzlin (1999). Physician fees and procedure intensity: the case of cesarean delivery. *Journal of health economics 18*(4), 473–490.

Hackmann, M. and V. Pohl (2018). Patient vs. provider incentives in long term care. *Working Paper*.

Iizuka, T. (2007). Experts' agency problems: evidence from the prescription drug market in japan. *The Rand journal of economics 38*(3), 844–862.

Iizuka, T. (2012). Physician agency and adoption of generic pharmaceuticals. *American Economic Review 102*(6), 2826–58.

Imbens, G. W. (2015). Matching methods in practice: Three examples. *Journal of Human Resources 50*(2), 373–419.

Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

Johnson, E. M. and M. M. Rehavi (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy 8*(1), 115–141.

Kamenica, E. (2019). Bayesian persuasion and information design. *Annual Review of Economics 11*, 249–272.

Kamenica, E. and M. Gentzkow (2011). Bayesian persuasion. *American Economic Review 101*(6), 2590–2615.

Li, M., Z. Li, and C.-M. W. Yip (2022). Informal payments and patients' perceptions of the physician agency problem: Evidence from rural china. *Social Science & Medicine 298*, 114853.

Lin, X. and C. Liu (2022). Credible persuasion. *arXiv preprint arXiv:2205.03495*.

Manning, W. G., J. P. Newhouse, N. Duan, E. B. Keeler, and A. Leibowitz (1987). Health insurance and the demand for medical care: evidence from a randomized experiment. *The American economic review*, 251–277.

McGuire, T. G. and M. V. Pauly (1991). Physician response to fee changes with multiple payers. *Journal of health economics 10*(4), 385–410.

Newhouse, J. P. et al. (1993). *Free for all?: lessons from the RAND health insurance experiment*. Harvard University Press.

Nguyen, A. and T. Y. Tan (2021). Bayesian persuasion with costly messages. *Journal of Economic Theory*, 105212.

Pauly, M. V. (2009). *Doctors and their workshops: Economic models of physician behavior*. University of Chicago Press.

Ran, L.-m., K.-j. Luo, Y.-c. Wu, L. Yao, and Y.-m. Feng (2013). An analysis of china's physician salary payment system. *Journal of Huazhong University of Science and Technology [Medical Sciences] 33*(2), 309–314.

Rowland, L. P. (1992). Surgical treatment of cervical spondylotic myelopathy: time for a controlled trial. *Neurology 42*(1), 5–5.

Sacks, D. W. (2018). Why do hmos spend less? patient selection, physician price sensitivity, and prices. *Journal of Public Economics 168*, 146–161.

Schnell, M. (2017). Physician behavior in the presence of a secondary market: The case of prescription opioids. Technical report, Mimeo.

Vatter, B. (2021). Quality disclosure and regulation: Scoring design in medicare advantage.

Wang, T., H. Wang, H.-D. A. Sen Liu, H. Liu, and W.-Y. Ding (2016). Anterior cervical discectomy and fusion versus anterior cervical corpectomy and fusion in multilevel cervical spondylotic myelopathy: a meta-analysis. *Medicine 95*(49).

Yip, W. C. (1998). Physician response to medicare fee reductions: changes in the volume of coronary artery bypass graft (cabg) surgeries in the medicare and private sectors. *Journal of* *health economics 17*(6), 675–699.

Online Appendices

A Model Details

A.1 Show Cutoff Strategy (Almost Everywhere) Optimal

Prove by contradiction. Suppose that the physician's optimal strategy is not the cutoff almost everywhere, as specified in (6). Then it must be the case that the physician recommends surgery with a probability strictly in $(0, 1)$ for some ξ_i with positive measure, and/or for some positive measure recommends surgery with probability 1 below some point and drugs with probability 1 above that point. Mathematically, I restrict my statements to intervals for simplicity, but the idea should carry beyond. These two situations can be summarized by the following: for some interval (c_1, c_3) of ξ_i, $-\infty \leq c_1 < c_3 \leq \infty$, the physician's recommendation strategy $\sigma_i(\xi_i)$ exhibits the following patterns: for $c_1 < c_2 < c_3$,

$$\sigma_i(\xi_i) = \begin{cases} p_1 \in (0,1] & \text{if } \xi_i \in (c_1, c_2] \\ p_2 \in [0,1) & \text{if } \xi_i \in (c_2, c_3) \end{cases} \tag{A.1}$$

Suppose that physician's optimal strategy is $\sigma_i(\xi_i)$ with the pattern shown in (A.1), and for $\xi_i \notin (c_1, c_3)$ physician gets expected utility of $\bar{M}$. Under this strategy, the physician's expected payoff is then

$$\begin{aligned} \mathbb{E}M^* &= \int_{c_1}^{c_2} p_1\big(\xi - \kappa_i + F_i\big)\, dG(\xi) + \int_{c_2}^{c_3} p_2\big(\xi - \kappa_i + F_i\big)\, dG(\xi) + \bar{M} \\ &= \int_{c_1}^{c_2} p_1\xi\, dG(\xi) + \int_{c_2}^{c_3} p_2\xi\, dG(\xi) + \bar{M} \\ &+ \underbrace{\big[p_1\big(G(c_2) - G(c_1)\big) + p_2\big(G(c_3) - G(c_2)\big)\big]}_{\equiv \mathbb{P}} \cdot (-\kappa_i + F_i) \end{aligned} \tag{A.2}$$

Also, patient's obedience constraints are satisfied:

$$\begin{aligned} &\mathbb{E}\big(\xi_i | \xi_i \in \{\xi : \sigma_i(\xi) > 0\}\big) - \kappa_i \geq 0 \\ &\mathbb{E}\big(\xi_i | \xi_i \in \{\xi : \sigma_i(\xi) < 1\}\big) - \kappa_i \leq 0 \end{aligned} \tag{A.3}$$

There must exist some $\tilde{c} \in (c_1, c_2)$, and $p^* > p_2$, such that

$$
\begin{aligned}
&\mathbb{P} = \big[p_1\big(G(c_2) - G(\tilde{c})\big) + p^*\big(G(c_3) - G(c_2)\big)\big] \equiv \mathbb{P}^{'} \\
\iff \quad & p_1(G(\tilde{c}) - G(c_1)) = (p^* - p_2)(G(c_3) - G(c_2))
\end{aligned} \tag{A.4}
$$

The next is to show that a strategy $\sigma_i^{'}(\xi_i)$ that moves equal probability mass from (c_1, c_2) to (c_2, c_3) gives the physician a strictly higher payoff than $\mathbb{E}M^*$, while the obedience constraints are still satisfied. The only difference between $\sigma_i^{'}(\xi_i)$ and $\sigma_i(\xi_i)$ is that for $\xi_i \in (c_1, c_3)$,

$$
\sigma_i^{'}(\xi_i) = \begin{cases} 0 & \text{if } \xi_i \in (c_1, \tilde{c}] \\ p_1 \in (0, 1] & \text{if } \xi_i \in (\tilde{c}, c_2] \\ p^* \in (0, 1] & \text{if } \xi_i \in (c_2, c_3). \end{cases} \tag{A.5}
$$

First, note that since $c_1 < \tilde{c}$ and $p^* > p_2$, the obedience constraints are still satisfied, and patients are even more willing to follow:

$$
\begin{aligned}
&\mathbb{E}\big(\xi_i|\xi_i \in \{\xi : \sigma_i^{'}(\xi) > 0\}\big) - \kappa_i > \mathbb{E}\big(\xi_i|\xi_i \in \{\xi : \sigma_i(\xi) > 0\}\big) - \kappa_i \geq 0 \\
&\mathbb{E}\big(\xi_i|\xi_i \in \{\xi : \sigma_i^{'}(\xi) < 1\}\big) - \kappa_i < \mathbb{E}\big(\xi_i|\xi_i \in \{\xi : \sigma_i(\xi) < 1\}\big) - \kappa_i \leq 0
\end{aligned} \tag{A.6}
$$

Based on $c_1 < \tilde{c} < c_2$ and (A.4),

$$
\begin{aligned}
&\int_{c_1}^{c_2} p_1 \xi \, dG(\xi) + \int_{c_2}^{c_3} p_2 \xi \, dG(\xi) \\
&= p_1(G(c_2) - G(c_1)) \cdot \mathbb{E}(\xi_i|c_1 < \xi_i \leq c_2) + p_2(G(c_3) - G(c_2)) \cdot \mathbb{E}(\xi_i|c_2 < \xi_i < c_3) \\
&< p_1(G(c_2) - G(c_1) \cdot \mathbb{E}(\xi_i|\tilde{c} < \xi_i \leq c_2) + p_2(G(c_3) - G(c_2)) \cdot \mathbb{E}(\xi_i|c_2 < \xi_i < c_3) \\
&= p_1(G(c_2) - G(\tilde{c})) \cdot \mathbb{E}(\xi_i|\tilde{c} < \xi_i \leq c_2) + p^*(G(c_3) - G(c_2)) \cdot \mathbb{E}(\xi_i|c_2 < \xi_i < c_3) \\
&+ p_1(G(\tilde{c}) - G(c_1)) \cdot \mathbb{E}(\xi_i|\tilde{c} < \xi_i \leq c_2) - (p^* - p_2)(G(c_3) - G(c_2)) \cdot \mathbb{E}(\xi_i|c_2 < \xi_i < c_3) \\
&= p_1(G(c_2) - G(\tilde{c})) \cdot \mathbb{E}(\xi_i|\tilde{c} < \xi_i \leq c_2) + p^*(G(c_3) - G(c_2)) \cdot \mathbb{E}(\xi_i|c_2 < \xi_i < c_3) \\
&+ \underbrace{p_1(G(\tilde{c}) - G(c_1)) \cdot \big[\mathbb{E}(\xi_i|\tilde{c} < \xi_i \leq c_2) - \mathbb{E}(\xi_i|c_2 < \xi_i < c_3)\big]}_{<0} \\
&< p_1(G(c_2) - G(\tilde{c})) \cdot \mathbb{E}(\xi_i|\tilde{c} < \xi_i \leq c_2) + p^*(G(c_3) - G(c_2)) \cdot \mathbb{E}(\xi_i|c_2 < \xi_i < c_3)
\end{aligned}
$$

$$= \int_{\tilde{c}}^{c_2} p_1 \xi \, dG(\xi) + \int_{c_2}^{c_3} p^* \xi \, dG(\xi) \tag{A.7}$$

As a result,

$$\begin{aligned}
\mathbb{E}M^* &= \int_{c_1}^{c_2} p_1 \xi \, dG(\xi) + \int_{c_2}^{c_3} p_2 \xi \, dG(\xi) + \mathbb{P} \cdot (-\kappa_i + F_i) + \bar{M} \\
&< \int_{\tilde{c}}^{c_2} p_1 \xi \, dG(\xi) + \int_{c_2}^{c_3} p^* \xi \, dG(\xi) + \mathbb{P} \cdot (-\kappa_i + F_i) + \bar{M} \\
&= \int_{\tilde{c}}^{c_2} p_1 \xi \, dG(\xi) + \int_{c_2}^{c_3} p^* \xi \, dG(\xi) + \mathbb{P}' \cdot (-\kappa_{\mathbb{i}} + \mathbb{F}_{\mathbb{i}}) + \bar{\mathbb{M}} \\
&= \int_{\tilde{c}}^{c_2} p_1 \big(\xi - \kappa_i + F_i\big) \, dG(\xi) + \int_{c_2}^{c_3} p^* \big(\xi - \kappa_i + F_i\big) \, dG(\xi) + \bar{M}
\end{aligned} \tag{A.8}$$

The last expression of (A.8) is the physician's expected payoff under strategy $\sigma'(\xi_i)$. However, this contradicts that $\mathbb{E}M^*$ is the maximum expected payoff under the optimal strategy.

A.2 Solve for Optimal Cutoff

Generic Form of Prior $G(\cdot)$. Karush–Kuhn–Tucker conditions provide the necessary conditions for physician's optimization. Let c_i^* be a local optimum. Let μ_1 and μ_2 be the KKT multipliers for the inequality constraints ① and ② in (8) separately, such that

$$\begin{cases}
-c_i^* \cdot g(c_i^*) - g(c_i^*) \cdot (-\kappa_i + F_i) + \mu_1 \dfrac{\partial\big(\mathbb{E}(\xi_i|\xi_i \geq c_i^*)\big)}{\partial c_i} - \mu_2 \dfrac{\partial\big(\mathbb{E}(\xi_i|\xi_i < c_i^*)\big)}{\partial c_i} = 0 & \text{(A.9)} \\
\mathbb{E}(\xi_i|\xi_i \geq c_i^*) - \kappa_i \geq 0 & \text{(A.10)} \\
\mathbb{E}(\xi_i|\xi_i < c_i^*) - \kappa_i \leq 0 & \text{(A.11)} \\
\mu_1 \cdot \big(\mathbb{E}(\xi_i|\xi_i \geq c_i^*) - \kappa_i\big) = 0 & \text{(A.12)} \\
\mu_2 \cdot \big(\mathbb{E}(\xi_i|\xi_i < c_i^*) - \kappa_i\big) = 0 & \text{(A.13)} \\
\mu_1 \geq 0 & \text{(A.14)} \\
\mu_2 \geq 0 & \text{(A.15)}
\end{cases}$$

where equation (A.9) is stationarity condition. (A.10) (A.11) (A.14) and (A.15) are feasibility conditions. (A.12) and (A.13) are complementary slackness.

Case 1: (A.10) binding

$$
\begin{aligned}
&\mathbb{E}(\xi_i|\xi_i \geq c_i^*) = \kappa_i \\
\implies &\kappa_i = \frac{\int_{c_i^*}^{\infty} \xi g(\xi) d\xi}{1 - G(c_i^*)} \equiv h(c_i^*) \qquad\qquad \text{(monotonic)} \\
\implies &c_i^* = h^{-1}(\kappa_i) < \kappa_i
\end{aligned}
$$

When (A.10) holds with equality, c_i^* can be written as an increasing function of κ_i, denoted as $h^{-1}(\kappa_i)$. By (A.12), $\mu_1 > 0$. Also note that because $c_i^* < \kappa_i$, (A.11) does not bind. By (A.13), $\mu_2 = 0$.

Case 2: (A.11) binding

$$
\begin{aligned}
&\mathbb{E}(\xi_i|\xi_i < c_i^*) = \kappa_i \\
\implies &\kappa_i = \frac{\int_{-\infty}^{c_i^*} \xi g(\xi) d\xi}{G(c_i^*)} \equiv m(c_i^*) \qquad\qquad \text{(monotonic)} \\
\implies &c_i^* = m^{-1}(\kappa_i) > \kappa_i
\end{aligned}
$$

When (A.11) holds with equality, c_i^* can be written as an increasing function of κ_i, denoted as $m^{-1}(\kappa_i)$. By (A.13), $\mu_2 > 0$. Since $c_i^* > \kappa_i$, (A.10) must be non-binding, which means $\mu_1 = 0$.

Note that (A.10) and (A.11) can never be binding at the same time. So the last case to discuss is when they are both non-binding.

Case 3: (A.10) and (A.11) non-binding

$$
\mathbb{E}(\xi_i|\xi_i < c_i^*) < \kappa_i < \mathbb{E}(\xi_i|\xi_i \geq c_i^*)
$$

In this case, $\mu_1 = \mu_2 = 0$. By equation (A.9),

$$
\begin{aligned}
&- c_i^* \cdot g(c_i^*) - g(c_i^*) \cdot (-\kappa_i + F_i) = 0 \\
\implies &c_i^* = \kappa_i - F_i
\end{aligned}
$$

To sum up the three cases,

$$
c_i^* = \begin{cases} h^{-1}(\kappa_i) & \text{if } \ \mathbb{E}(\xi_i|\xi_i \geq \kappa_i - F_i) < \kappa_i \\ m^{-1}(\kappa_i) & \text{if } \ \mathbb{E}(\xi_i|\xi_i < \kappa_i - F_i) > \kappa_i \\ \kappa_i - F_i & \text{otherwise} \end{cases} .
$$

Since the empirical fact indicates that F_i is not negative, case 2 where $c_i^* = m^{-1}(\kappa_i)$ can be ruled out. When F_i is strictly positive, the optimal cutoff c_i^* can be written as

$$c_i^* = \max\{h^{-1}(\kappa_i), \kappa_i - F_i\} \tag{A.16}$$

where $h^{-1}(x)$ is the inverse function of $h(x) = \mathbb{E}(\xi_i|\xi_i \geq x) \equiv \frac{\int_x^\infty \xi g(\xi)d\xi}{1-G(x)}$.

When $F_i = 0$ (this could be achieved by letting $\beta = 0$ or $T_i^s = T_i^d$), $c_i^* = \kappa_i$, meaning that the physician's recommendation is without any bias. To integrate the case of $F_i = 0$ into (A.16), it requires $\kappa_i > h^{-1}(\kappa_i)$, or $h(\kappa_i) > \kappa_i$ given that $h(x)$ is increasing in x. It turns out $h(\kappa_i) > \kappa_i$ holds for all distributions $\boldsymbol{G}$ with thin right tails. To see this, note that

$$\begin{aligned}
h(\kappa_i) =& \frac{\int_{\kappa_i}^\infty \xi g(\xi)d\xi}{1 - G(\kappa_i)} \\
=& \frac{-\xi(1 - G(\xi))\big|_{\kappa_i}^\infty + \int_{\kappa_i}^\infty (1 - G(\xi))d\xi}{1 - G(\kappa_i)} \\
=& \frac{-\lim_{\xi\to\infty} \xi(1 - G(\xi))}{1 - G(\kappa_i)} + \kappa_i + \frac{\int_{\kappa_i}^\infty (1 - G(\xi))d\xi}{1 - G(\kappa_i)} \\
>& \frac{-\lim_{\xi\to\infty} \xi(1 - G(\xi))}{1 - G(\kappa_i)} + \kappa_i
\end{aligned} \tag{A.17}$$

$h(\kappa_i) > \kappa_i$ holds if $\lim_{\xi\to\infty} \xi(1-G(\xi)) = 0$, which implies that the right tail should be thin enough. With this thin tail assumption and $F_i \geq 0$, the optimal cutoff takes the form of (A.16).

Prior $\boldsymbol{G}(\cdot)$ as Standard Normal. $\Phi(\cdot)$ and $\phi(\cdot)$ denote the CDF and PDF of standard normal respectively. When $G \sim N(0, 1)$,

$$\begin{aligned}
h(x) &= \mathbb{E}(\xi_i|\xi_i \geq x) \\
&= \int_x^\infty \frac{\xi \cdot \phi(\xi)}{1 - \Phi(x)} d\xi \\
&= \frac{1}{1 - \Phi(x)} \cdot \int_x^\infty (-\phi'(\xi))d\xi \\
&= \frac{-\phi(\infty) + \phi(x)}{1 - \Phi(x)} \\
&= \frac{\phi(x)}{1 - \Phi(x)}
\end{aligned} \tag{A.18}$$

Note that $\frac{\phi(x)}{1-\Phi(x)} = \frac{\phi(-x)}{\Phi(-x)}$, which is the Inverse Mills ratio at $-x$. Denote the Inverse Mills ratio as $\lambda(.)$. The optimal cutoff is

$$c_i^* = \max\{-\lambda^{-1}(\kappa_i), \kappa_i - F_i\} \tag{A.19}$$

A.3 Prove Model Properties

Proof of Property 1. The derivative of c_i^* w.r.t F_i is:

$$\frac{\partial(c_i^*)}{\partial(F_i)} = \mathbb{1}\{h^{-1}(\kappa_i) < \kappa_i - F_i\} \cdot (-1) + \mathbb{1}\{h^{-1}(\kappa_i) \geq \kappa_i - F_i\} \cdot (0). \tag{A.20}$$

First, I show that $h^{-1}(\kappa_i) - (\kappa_i - F_i)$ increases in κ_i. Note that

$$\begin{aligned} h(x) =& \frac{\int_x^\infty \xi g(\xi) d\xi}{1 - G(x)} \\ =& \frac{-\xi(1-G(\xi))|_x^\infty + \int_x^\infty (1-G(\xi))d\xi}{1-G(x)} \\ =& x + \frac{\int_x^\infty (1-G(\xi))d\xi}{1-G(x)}. \end{aligned} \tag{A.21}$$

Therefore,

$$\begin{aligned} h^{'}(x) - 1 =& \frac{g(x)\int_x^\infty (1-G(\xi))d\xi}{(1-G(x))^2} - 1 \\ =& \frac{g(x)\int_x^\infty (1-G(\xi))d\xi - (1-G(x))^2}{(1-G(x))^2} \\ <& 0 \quad \text{iff} \int_x^\infty (1-G(\xi))d\xi \quad \text{log-concave.} \end{aligned} \tag{A.22}$$

Since $h^{'}(\kappa_i) > 0$, $h^{'}(\kappa_i) - 1 < 0$ is equivalent to $\frac{1}{h'(\kappa_i)} - 1 > 0$, or $h^{-1}(\kappa_i) - (\kappa_i - F_i)$ increasing in κ_i .

Therefore, when κ_i is small, or patients are more insured, the obedience constraint is less likely to bind and $\frac{\partial(c_i^*)}{\partial(F_i)} = -1$. When κ_i rises (when patients are less insured) to a point, the obedience constraint will bind and $\frac{\partial(c_i^*)}{\partial(F_i)} = 0$. In such cases, physicians do not respond to the change in F_i. Q.E.D.

Proof of Property 2. The derivative of c_i^* w.r.t ΔT_i is:

$$\frac{\partial(c_i^*)}{\partial(\Delta T_i)} = \mathbb{1}\{h^{-1}(\kappa_i) < \kappa_i - F_i\} \cdot (\frac{\partial k_i}{\partial \Delta T_i} - f^{'}(\Delta T_i)) + \mathbb{1}\{h^{-1}(\kappa_i) \geq \kappa_i - F_i\} \cdot \frac{\partial k_i}{\partial \Delta T_i} \cdot (\frac{1}{h^{'}(\kappa_i)}). \tag{A.23}$$

If $c_i^* = \kappa_i - F_i$, it is decreasing in ΔT_i as long as $\frac{\partial k_i}{\partial \Delta T_i} < f^{'}(\Delta T_i)$ (physicians' incentives dominate). As ΔT_i increases to the point when $\kappa_i - F_i$ drops below $h^{-1}(\kappa_i)$, $c_i^* = h^{-1}(\kappa_i)$, and it is increasing in ΔT_i since $h^{'}(\cdot) > 0$ and $\frac{\partial k_i}{\partial \Delta T_i} > 0$. Q.E.D.

Proof of Property 3. Taking the derivative of the distance between the patient's and physician's cutoffs w.r.t κ_i gives

$$\frac{\partial(\kappa_i - c_i^*)}{\partial \kappa_i} = \mathbb{1}\{h^{-1}(\kappa_i) \geq \kappa_i - F_i\} \cdot (1 - \frac{1}{h'(\kappa_i)}) + \mathbb{1}\{h^{-1}(\kappa_i) < \kappa_i - F_i\} \cdot 0. \tag{A.24}$$

(A.22) has already proved that $1 - \frac{1}{h'(\kappa_i)} < 0$. Therefore, for the constrained physicians, $\frac{\partial(\kappa_i - c_i^*)}{\partial \kappa_i} < 0$. Q.E.D.

Proof of Property 4. First, note that $c_i^* \leq \kappa_i$ with the equality holds if and only if $F_i = 0$ (Appendix A.2 includes this discussion). Second, recall that a patient under full information has his own (first best) cutoff: he prefers surgery if and only if $\xi_i > \kappa_i$. Q.E.D.

Property 5. *With incentives misalignment, if patient i becomes fully informed, there is a strictly expected utility gain for the patient and utility loss for the physician.*

Proof of Property 5. It is straightforward to show that granting patient full information makes the patient strictly better off. This is because a fully informed patient can always ignore the new information and set his own cutoff to what the physician employs under asymmetric information. According to Property 4, doing so is not optimal. Similarly, the physician is strictly worse off, because she can always choose to reveal accurate information ($c^* = \kappa_i$) to an uninformative patient. However, it is not optimal for her to do so with incentive misalignment ($c^* < \kappa_i$).

Mathematically, a patient changing from partial to full information will increase the optimal cutoff from $c_i^* = \max\{h^{-1}(\kappa_i), \kappa_i - F_i\}$ to $c_i^{*'} = \kappa_i$. The cutoff change will affect the recommendation for patient i with ξ_i such that $c_i^* \leq \xi_i < c_i^{*'}$. The expected change in patient utility is

$$\int_{\max\{h^{-1}(\kappa_i), \kappa_i - F_i\}}^{\kappa_i} (\kappa_i - \xi) \, d\, G(\xi). \tag{A.25}$$

Given the integral interval for ξ, the expected utility change for the patient is strictly positive under incentive misalignment. For the physician, she gains the patient's utility gain, but loses some utility due to remuneration loss, the expected change in physician utility is

$$\int_{\max\{h^{-1}(\kappa_i),\kappa_i-F_i\}}^{\kappa_i} (\kappa_i - \xi - F_i)\, d\, G(\xi). \tag{A.26}$$

Note that for ξ within this integral interval,

$$-F_i < \kappa_i - \xi - F_i \leq \min\{-h^{-1}(\kappa_i), F_i - \kappa_i\} + \kappa_i - F_i \leq 0. \tag{A.27}$$

Q.E.D.

A.4 Relation to the Leading Example in Kamenica and Gentzkow (2011)

To make connections to the canonical example of a prosecutor persuading a judge, I first show that after partitioning the continuous unknown state to binary, my model delivers some optimal signal properties as in KG. I define a binary state: $\xi_i \geq \kappa_i$ or $\xi_i < \kappa_i$. This means the two states are "surgery better or equal" ($U_i^s \geq U_i^d$) and "surgery worse" ($U_i^s < U_i^d$). Following KG, the physician chooses a signal structure with two probabilities: the probability of recommending surgery conditional on surgery giving patient higher utility, denoted as p_1, and the probability of recommending surgery conditional on surgery giving patient lower utility, denoted as p_2.

$$\begin{aligned}
p_1 &\equiv Pr(r_i = s | U_i^s \geq U_i^d) \\
&= Pr(\xi_i \geq c_i^* | \xi_i \geq \kappa_i) \\
&= 1 \\
\\
p_2 &\equiv Pr(r_i = s | U_i^s < U_i^d) \\
&= Pr(\xi_i \geq c_i^* | \xi_i < \kappa_i) \\
&= \frac{Pr(c_i^* \leq \xi_i < \kappa_i)}{Pr(\xi_i < \kappa_i)} \\
&= \frac{G(\kappa_i) - G(c_i^*)}{G(\kappa_i)} \in [0, 1)
\end{aligned} \tag{A.28}$$

Note that $p_2 = 0$ corresponds to $c_i^* = \kappa_i$, meaning physician's and patient's incentives are perfectly aligned, and the information is accurate. In a world where incentive misalignment exists, there are two properties of the signal structure consistent with KG. First, the signal receiver (the patient) is only certain about the state when he chooses the sender's (physician's) least-preferred action (d). This is because while $p_1 = 1$ implies that the physician will recommend s for sure if s indeed gives higher utility to the patient, $p_2 > 0$ suggests that the physician recommends surgery even to some patients who benefit more from drugs. As a result, a patient receiving a surgery recommendation is never certain about the state. Second, under the optimal signal $c_i^* = h^{-1}(\kappa_i)$ (corner solution in equation (9)), the patient is exactly indifferent between s and d when choosing s. That is, the constrained physician will "persuade to the maximum".

However, unlike KG, the obedience constraints can be non-binding for some physicians under the optimal strategy. In my case, the physician is altruistic to the patient, and she must weigh the patient's benefit against her remuneration interest. As a result, persuading to the maximum may not be optimal.

B Insurance Details

An insurance contract is structured in the following way: at each visit, the patient pays the medical charge entirely out-of-pocket until reaching the deductible amount[34]. After reaching the deductible amount, the patient then pays a fixed percentage (i.e., coinsurance rate) of the covered portion of care for subsequent medical services, while his insurance plan pays the rest. Within one plan-year, however, once the covered amount exceeds an allowed reimbursement ceiling, the patient pays 100 percent of all his medical expenses for the remainder of that year.

I compare the generosity of the two government-provided insurance plans by plotting the out-of-pocket price, given the total charge of medical treatment under each plan. To display the annual coinsurance limit in the graph, I consider the situation when a patient visits a hospital for the first time in a plan-year.

Note that since the financial characteristics could vary within an insurance type based on hospital tiers, plan-year, and patient retirement status, I take the simple average of the coinsurance rate across hospital tiers and plan-years for each insurance type for simplicity. As shown in Figure B.1, based on the averaged financial characteristics, the deductible for urban employment insurance is \$46 lower than that for the resident insurance. For a total price higher than the deductible, the coinsurance rate is 10 percent lower, and the allowed reimbursement ceiling is \$8,572 higher for the urban employment insurance. Due to these differences, urban employment insurance has a weakly lower marginal out-of-pocket price.

[34]The deductible resets at every visit. This differs from the U.S., where the deductible resets yearly.

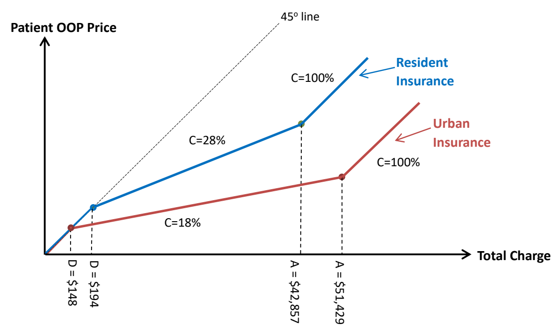


Figure B.1: Insurance Financial Characteristics and Patient Out-of-Pocket Price

Note: *The figure plots a patient's out-of-pocket price against the total charges for medical treatment for each insurance plan based on its deductible (D), coinsurance rate (C), and allowed reimbursement ceiling (A), assuming this is the first time a patient uses insurance in that plan year. Since the financial characteristics are different across hospital tiers (primary, secondary, and tertiary) and plan-years within each insurance type, I use the simple average of deductibles and coinsurance rates across hospital tiers and years. Plotting the average is without loss of generality because the urban employment insurance always offers a lower marginal out-of-pocket price conditional on hospital tier in a given year.*

C Hospital Characteristics Before Policy Change

Table C.1: Summary Statistics by Hospitals: Pre-policy

	All treatment hos (1)	Treatment hos 1 (2)	Treatment hos 2 (3)	All control hos (4)	Control hos 1 (5)	Control hos 2 (6)
Demographics						
Age	54.44	56.83	53.51	53.88	50.16	55.35
	(13.57)	(14.46)	(11.91)	(13.13)	(15.19)	(11.21)
Gender (male=1)	0.39	0.29	0.30	0.36	0.47	0.43
Insurance (urban=1)	0.59	0.54	0.62	0.49	0.58	0.56
Coinsurance rate	0.24	0.24	0.29	0.18	0.16	0.19
	(0.13)	(0.15)	(0.18)	(0.06)	(0.03)	(0.07)
Subsidy (yes=1)	0.26	0.22	0.18	0.23	0.24	0.29
Comorbidity (yes=1)	0.25	0.17	0.28	0.38	0.32	0.37
Treatment Outcomes						
Treatment Mode (surgery =1)	0.04	0.03	0.04	0.07	0.11	0.06
Treatment Charge ($ in thousands)	1.62	1.67	1.76	1.40	1.59	1.13
	(1.20)	(1.27)	(1.36)	(0.91)	(0.80)	(0.51)
-Surgical	4.33	4.13	4.16	2.32	3.41	2.87
	(3.18)	(1.07)	(1.52)	(2.23)	(1.22)	(1.60)
-Non-surgical	1.50	1.60	1.66	1.33	1.37	1.02
	(1.10)	(0.79)	(1.26)	(0.67)	(0.81)	(0.51)
Readmission	0.01	0.00	0.01	0.01	0.02	0.01
Observations	1,351	562	203	3,859	579	141

Note: *Means and standard deviations (in parentheses) are displayed. Columns (1) and (4) show hospital characteristics by treatment and control hospitals. Columns (2), (3), (5), and (6) display a specific hospital. These hospitals were selected to illustrate the coinsurance rate variation after conditioning on hospital fixed effects.*

D Fee Differential Between Surgical and Non-surgical Treatment

Each treatment can be seen as a "bundle of categorical procedures", with category including medication, tests, therapies, anesthesia, in-hospital nursing, or/and surgeries. At time t, $t \in \{0, 1\}$, for procedure r of category c in treatment j, $j \in \{s, d\}$, denote its averaged quantity as q^j_{rct} and price as p_{rct} (the price is the same across treatments). Then holding the quantities at its pre-policy level, the total fee for treatment j at time t is

$$P^j_t = \sum_c \sum_{r \in c} p_{rct} \cdot q^j_{rc0}. \tag{D.1}$$

The average fee in each category before the policy change in treated hospitals, i.e., $\sum_{r\in c} p_{rc0} \cdot q^j_{rc0}$ for each c, can be calculated from the data. Holding the bundles constant, the fee in category c under the new price menu is $\sum_{r\in c} p_{rc1} \cdot q^j_{rc0}$. This object can also be calculated using the price changes in each category.

Therefore, fees for pre-policy, P_0^s and P_0^d, and fees for post-policy, P_1^s and P_1^d, can be derived. As a result, the fee differential at time t for surgical treatment is $P_t^s - P_t^d$. The percentage change in fee differential over time is then $\frac{(P_1^s - P_1^d) - (P_0^s - P_0^d)}{P_0^s - P_0^d}$. Table D.2 illustrates how the price changes in each category translate into the overall charge changes for each treatment option.

Table D.2: Total Charge ($) in Treated Hospitals

		Surgical Treatment		Non-surgical Treatment		Differential		% Change in Differential
Category	Menu Version	Old	New	Old	New	Old	New	
Surgical Procedures		1,284	1,670	0	0			
Drugs		311	271	438	381			
Bed, Routine Check, and General Nursing		297	445	157	235			
Others		2,439	2,439	909	909			
Total		4,331	4,825	1,504	1,525	2,827	3,299	17%

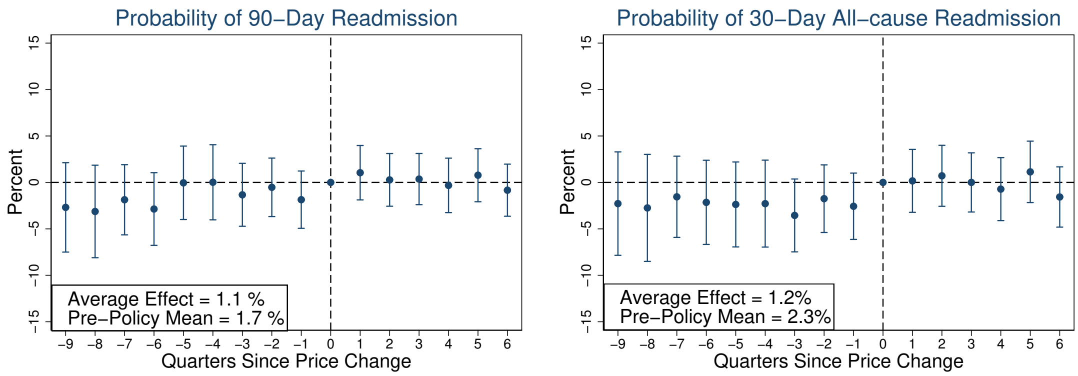
Note: *Other categories include medical materials, lab tests, anesthesia, and therapies. The charge within a category under the old menu is the average charge in treated hospitals before the policy; i.e., the old prices multiplied by the old average quantities. Charges under the new menu is the predicted charge based on the price changes specified in the policy, assuming the same quantities of procedures/goods in each category as before; i.e., the new prices times the old average quantities. For example, for surgical procedures, the new charge equals the old charge multiplied by 130 percent. For drugs, the new charge equals the old charge divided by 115 percent. For hospital bed, physician routine check and nursing, the new charge is the old charge multiplied by 150 percent.*

E Details in Difference-in-differences Analysis

E.1 Event Study with Alternative Definitions of Readmission

Replacing 30-day Neck arthritis specific readmission with 90-day Neck arthritis specific readmission and 30-day all-cause readmission respectively in equation (12), Figure E.2 show that the policy change did not significantly impact the likelihood of patient readmission under alternative definitions of readmission.

Figure E.2: Effect of Policy Change on Readmission



Note: *The figure shows the effect of the policy on readmissions for Cervical Spondylosis. It plots regression coefficients for the dynamic effects and their 95 percent confidence intervals in equation (12), using 90-day neck arthritis specific readmission dummy (left) and 30-day all-cause readmission dummy (right) as the dependent variable respectively. Effects are normalized to the end of the quarter just before the policy. The number of observations is 10,596. Standard errors are clustered at the hospital level.*

E.2 Event Study with a Linear Trend Difference

In one specification, I include the treatment group-specific time trend, in place of the interactions between the lagged time period dummies and the treatment group indicator. The dynamic treatment effects and confidence intervals are presented in red in Figure E.3. I refer to this specification as "without lags" since all the lagged treatment group-specific time fixed effects are dropped. In the other specification, I add a treatment group-specific linear trend on top of the interactions in the pre-period. Here, the pre-trend is identified by the first and last quarters in the pre-period, as I omit the treatment group-specific time fixed effects for these two quarters to avoid multi-linearity. I refer to this specification as "with lags." The corresponding treatment effects are presented in navy in Figure E.3.

From the figure, the treatment effect estimates are not sensitive to the alternative specifications; both specifications replicated the treatment effects of the main specification in equation (12) which restricts the pre-trend to zero. Besides this, a t-test of the linear pre-trend fails to reject the null.

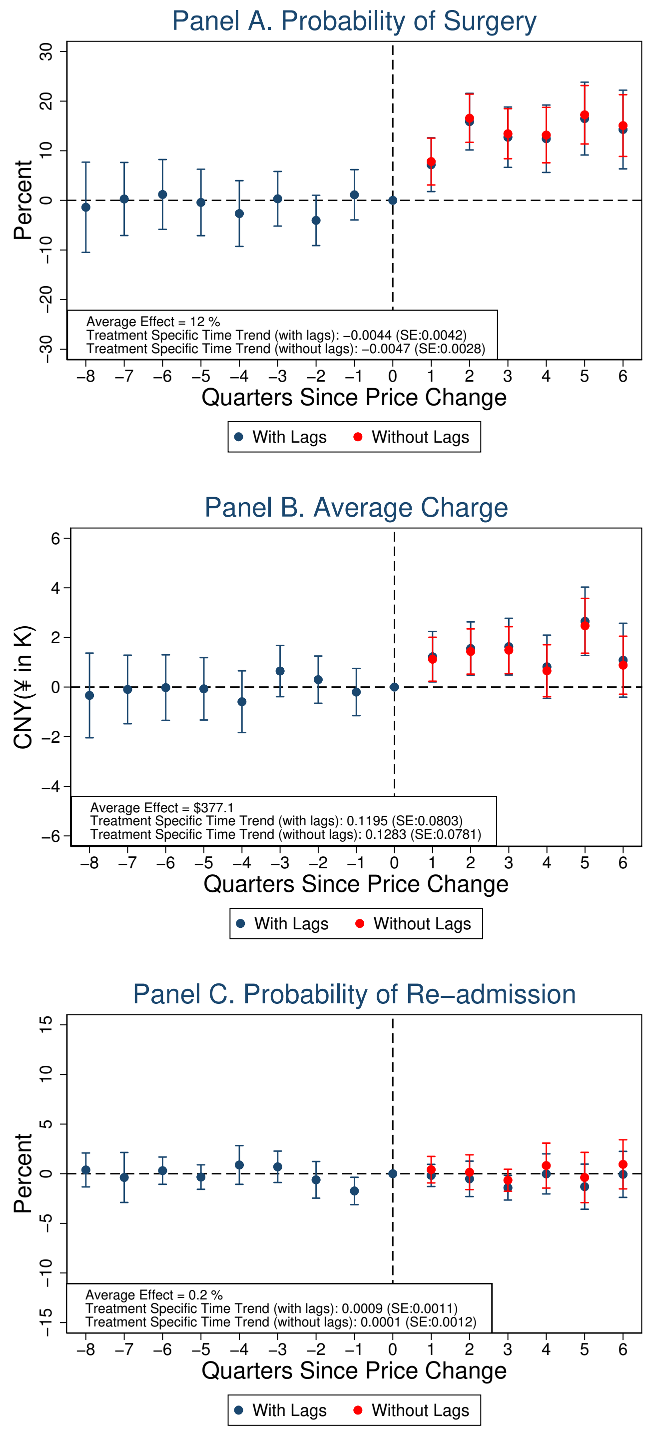


Figure E.3: Effect of Price Change on Health Care Utilization and Outcome: Event Study

Note: *The figure shows the policy's effect on health care utilization and outcome measures for neck arthritis. It plots regression coefficients for the policy change and their 95 percent confidence intervals by adding a treatment-specific linear time trend ("with lags," in navy) or replacing the lagged treatment-specific time fixed effects with a treatment specific trend ("without lags," in red) in equation (12). Effects are normalized to the end of the quarter just before the policy. The number of observations is 10,596. Standard errors are clustered at the hospital level.*

E.3 Rule Out Patients Sorting

To rule out possible patient sorting based on their observed characteristics, I test whether the patients' composition changed in the treated hospitals due to the policy. I estimate equation (12) by using patient age, gender, insurance type, indicator for subsidy, and indicator for comorbidity as the outcome measures respectively. The coefficients on the policy change are displayed in Figure E.4, which indicates no significant compositional change. In addition to the dynamic effects, I run regression (12) by replacing the six post-policy quarter indicators with a single post-policy indicator to show the average effects. As shown in Table E.3, the average changes in those characteristics are not significant due to the policy change at the 5 percent significance level.[35]

Table E.3: Average Effects of Price Change on Patient Composition

	Age	Gender (male=1)	Insurance Type (more =1)	Subsidy (yes =1)	Comorbidity (yes=1)
$D_h \cdot Post_t$	-1.012	0.006	0.047	-0.004	-0.042
	(0.988)	(0.040)	(0.023)	(0.073)	(0.036)

Note: *The table shows the regression coefficients for the policy change in equation (12) by replacing the six post-policy quarter indicators with a single post-policy indicator. The number of observations is 10,596. Standard errors are in parentheses and clustered at the hospital level.*

To rule out sorting based on unobservables due to patient price sensitivity, I discuss several inconsistencies between my data and the implications of patient selection. First, if patients who are inclined to surgery select disproportionally into control hospitals (where surgical treatment became relatively cheaper after the policy change), I should see a higher surgery rate in the control hospitals (note that physicians' incentives in control hospitals did not change around the treatment time). To the extent that patients inclined toward drugs might disproportionally select into treated hospitals, I expect an even higher surgery rate in the control hospitals after the policy change. I therefore estimate the following equation using the sample in control hospitals:

$$Sur_{iht} = \alpha_h + \sum_{k=-9}^{-1} \beta_k \cdot Q_k + \sum_{k=1}^{6} \beta_k \cdot Q_k + \epsilon_{iht}. \tag{E.1}$$

Sur_{iht} is a dummy equal to 1 if patient i in control hospital h took surgical treatment. α_h is the hospital fixed effects, and Q_k is an indicator variable for the k^{th} quarter-year relative to the time of the policy. As shown in Figure E.5, the surgery rates did not change significantly after the policy.

[35] The coefficient for the insurance type approaches the borderline of significance (p=0.057).

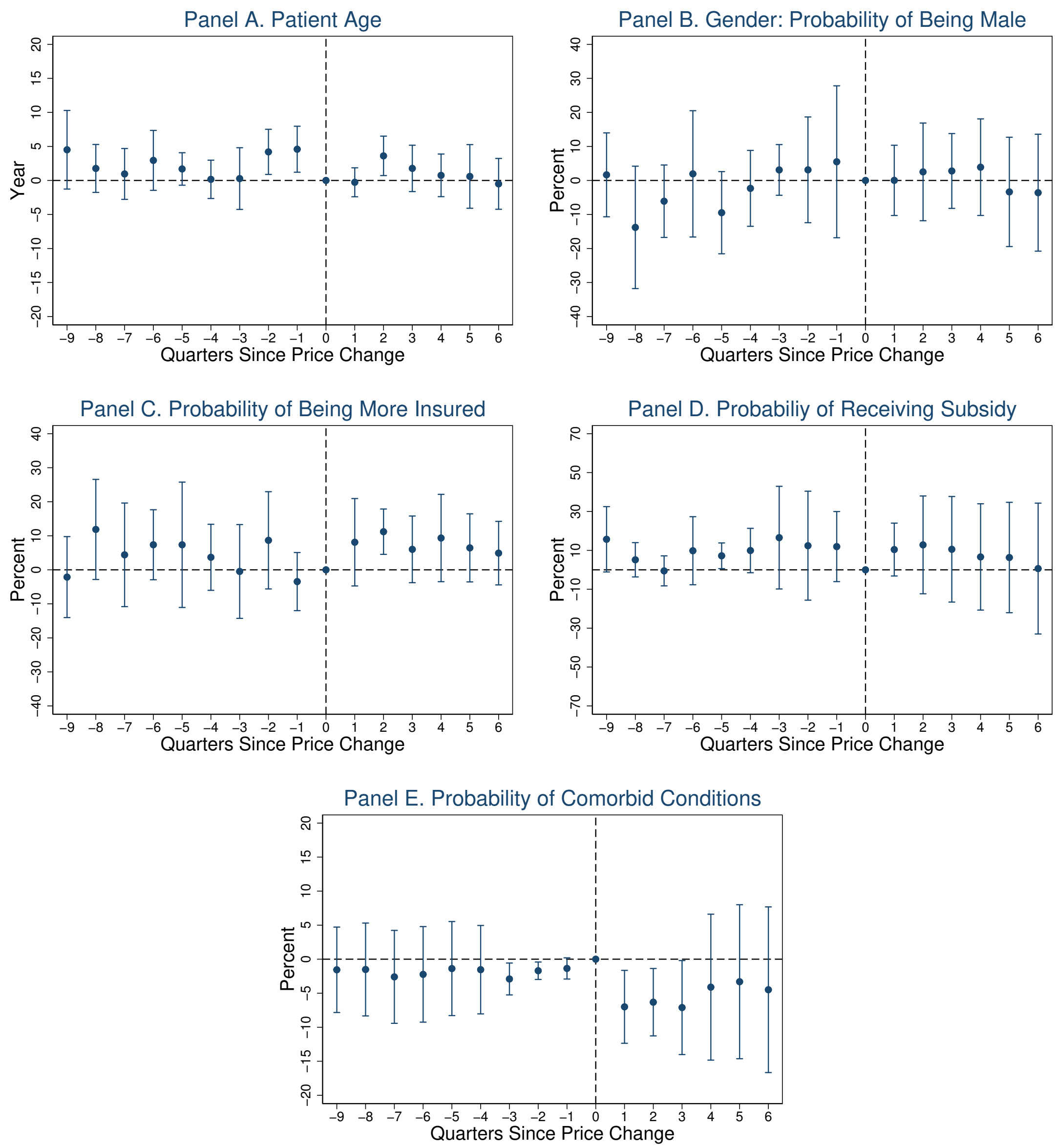


Figure E.4: Effect of Price Change on Patients Composition

Note: *The figures show the effects of the policy on patient characteristics in treated hospitals. Each subfigure plots regression coefficients for the policy change and their 95 percent confidence intervals from equation (12), using age (panel A), gender (panel B), insurance type (panel C), and subsidy indicator (panel D), and comorbidity indicator (panel E) as the dependent variables. Effects are normalized to the end of the quarter just prior to the policy. The number of observations is 10,596. Standard errors are clustered at the hospital level.*

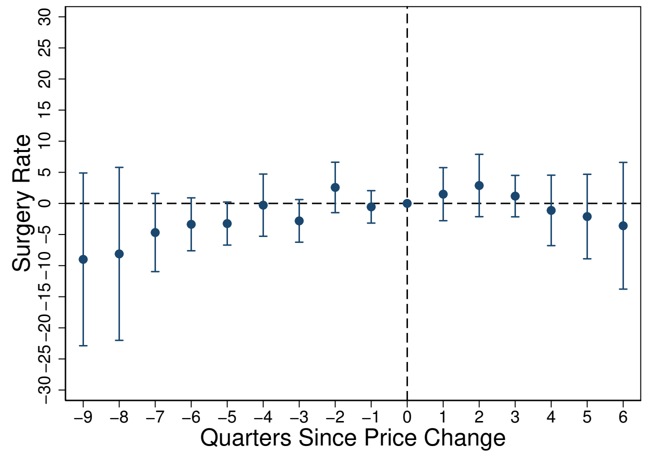


Figure E.5: Surgery Rate in Control Hospitals

Note: *The figure shows the quarterly mean of surgery rate in control hospitals. It plots quarter dummy coefficients and their 95 percent confidence intervals from equation (E.1). Outcomes are normalized to the end of the quarter just prior to the policy. The number of observations is 10,596. Standard errors are clustered at the hospital level.*

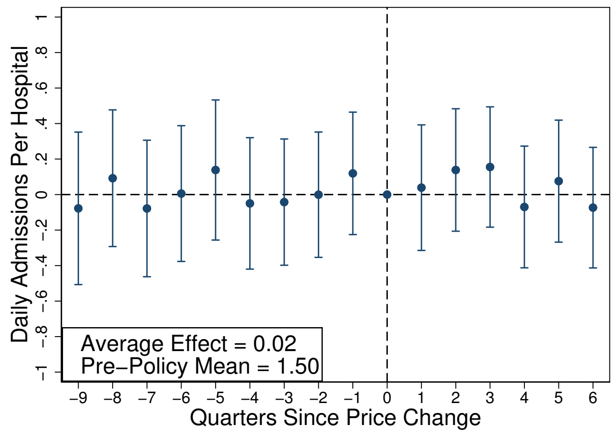
Second, if there were patient selection behavior, total volume of patients in treated and control hospitals should change in an opposite way. I report the total number of patients in treated and control hospitals before and after the policy in Table E.4. The market share of treated hospitals before the policy was 27 percent, and it was 28 percent afterward. These two numbers are fairly close, indicating that a substantial selection is very unlikely. To strengthen the argument that there is no differential trend in admission volume between control and treated hospitals, I run regression (12) with daily admission volume as the outcome variable. The results plotted in Figure E.6 show that the policy did not affect the total patient volume in treated hospitals.

Table E.4: Admission Volume for Neck Arthritis

	Before		After	
	Admission Volume	Market Share	Admission Volume	Market Share
Control Hospitals	3,796	72.9%	3875	71.9%
Treated Hospitals	1,414	27.1%	1511	28.1%
Total	5,210	100.0%	5386	100.0%

Third, following the same logic, for any given diagnosis for which surgery is the only treatment option, the selection model should have many more patients who switched hospitals after the policy change. I therefore expect to see much larger disparities in the market share across time for each hospital group. In my data, thus, I now examine the change in patient volumes

Figure E.6: Policy Effect on Patient Volume for Neck Arthritis



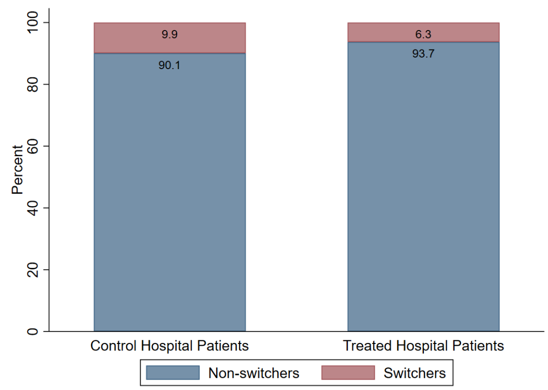
for patients with age-related cataracts, over 95 percent of whom underwent surgical procedures. In Table E.5, the market share of this diagnosis for the treated hospitals hardly changed around the time of the policy change (from 13.6 percent before to 13.1 percent after), thus implying minimal price-based selection if indeed patients selected on price at all.

Table E.5: Admission Volume for Age-related Cataracts

	Before		After	
	Admission Volume	Market Share	Admission Volume	Market Share
Control Hospitals	4,083	86.4%	4229	86.9%
Treated Hospitals	645	13.6%	640	13.1%
Total	4,728	100.0%	4869	100.0%

Lastly, at a more aggregate level, the selection model implies that a significant amount of price-sensitive patients should switch between control and treated hospitals due to the policy change, based on their tendency to opt for surgical and non-surgical treatments, *regardless of diagnoses*. However, evidence from the data contradicts this implication. In particular, the whole sample, including 374,067 patients with various diagnoses, suggests that patients who always went to control hospitals before the policy change continued to seek treatment in control hospitals afterward, with rare exceptions (Figure E.7). On the other hand, patients who only went to treated hospitals before the policy change nearly all stuck to treated hospitals afterward as well. In other words, patients are largely "non-switchers" after the policy change.

Figure E.7: Patient Hospital Switching Behaviors After Policy Change



Note: *The figure shows the proportion of patients switching to treated hospitals among those who always went to control hospitals before the policy (control hospital patients, N=15,797, the left bar) and the proportion of patients switching to control hospitals among those who always went to treated hospitals before the policy (treated hospital patients, N=29,629, the right bar).*

E.4 Balance Check by Patient Insurance Types Using External Data

I use the most recent wave (from 2020) of a high-quality longitudinal survey in China, the China Family Panel Studies (CFPS)[36], with around 20,000 nationwide representative observations on insurance type and various demographics. I regress each variable that affects a patient's surgery propensity but is unobserved in my data on the insurance type dummy. I report the results both for the regression that uses the whole sample and the regression that is conditional on the sample with urban residence, since my data cover only individuals living in urban areas regardless of their insurance type.

The results in Table E.6 indicate that, whether conditional on urban residence or not, variables that proxy for health conditions show no significant differences by insurance type, except for "recent discomfort" and "age". Unreported normalized differences indicate that differences in these two variables are not economically significant. In addition to health conditions, income can be an important factor in determining surgery decisions. Since the CFPS does not contain comprehensive income information, I use data from the Chinese Longitudinal Healthy Longevity Survey (CLHLS) to analyze the balance of income.[37] As most people would have

[36]The official website for the CFPS is https://www.isss.pku.edu.cn/cfps/en/.

[37]Zeng, Yi, Vaupel, James, Xiao, Zhenyu, Liu, Yuzhi, and Zhang, Chunyuan. Chinese Longitudinal Healthy Longevity Survey (CLHLS), 1998-2014. Inter-university Consortium for Political and Social Research [distributor], 2017-04-11. https://doi.org/10.3886/ICPSR36692.v1. According to this data, the difference in health status, chronic diseases, drinking, and smoking habits are similar between the two insurance types, consistent with the finding from the CFPS data. However, I choose not to use the CLHLS as the primary data source because the CLHLS focuses on an elderly population.

imagined, in the full sample, the more insured have a significantly higher income level, since they possess urban registration and are employed. However, once I examine both insurance types who live urban areas, as in my own data, there is no significant difference in income, and the income level seems to be even slightly higher for the less insured.

Table E.6: Balance Check by Insurance Type: External Data

		Unconditional		Conditional on Urban Residence	
Variables	**Explanation**	**Diff btw Insurance types**	**P-value**	**Diff btw Insurance types**	**P-value**
Health Conditions					
Self-reported Health Status	Ordinal, 1 being excellent health and 5 poor health	0.001	(0.95)	0.026	(0.30)
Chronic Disease	Dummy, whether diagnosed within past 6 months	-0.002	(0.81)	0.005	(0.49)
Recent Discomfort	Dummy, whether felt discomfort during last 2 weeks	-0.030	(0.00)	-0.017	(0.08)
Health Compared to Last Year	Ordinal, 1 being better; 3 no change; 5 worse	-0.006	(0.79)	-0.001	(0.97)
Known Illness Last Year	Dummy, whether has known illness during the past year	0.051	(0.00)	0.049	(0.00)
Smoking	Dummy, whether smoked cigarettes last month	-0.008	(0.31)	0.009	(0.32)
Alcohol Consumption	Dummy, whether drank alcohol 3 times per week last month	-0.001	(0.84)	-0.003	(0.66)
Age	Continuous, in years	-1.724	(0.00)	-0.270	(0.44)
Income					
Annual Income (Data from CLHLS)		17,700.400	(0.00)	-2,573.449	(0.28)
Trust in Physicians					
Rating on Medical Expertise	Ordinal, 1 being very poor and 5 very good	0.018	(0.25)	0.025	(0.18)
Risk Altitude					
Rating on being Adventurous	Dummy, self-evaluated	0.007	(0.55)	-0.003	(0.81)
Awareness of Policy Change/Incentives					
Importance of News	Ordinal, from 1 to 5 increasing importance of broadcast news	-0.037	(0.14)	0.051	(0.08)
Healthcare Consumption					
Total Annual Medical OOP Expenses	Continuous, in CNY	-25.020	(0.88)	230.100	(0.29)
Total Annual Medical Expenses	Continuous, in CNY	522.600	(0.05)	888.800	(0.01)

Note: *This table reports the difference in means for each variable by insurance type (more insured minus less insured) and its P-value, using a full sample and sub-sample with the urban residence. Data except income come from China Family Panel Studies, with 19,625 observations in the full sample and 9,899 in the sub-sample. Income data are from the Chinese Longitudinal Healthy Longevity Survey (CLHLS), with 5,814 observations in the full sample and 713 in the sub-sample.*

Next, I use patients' rating of their hospitals' expertise as a proxy for patients' trust in physicians, patients' evaluation of whether they are adventurous as a proxy for their risk aversion, and patients' rating on the importance of news in daily life as a reflection of their awareness of physicians' incentives and hospital policy changes. The estimates imply that the insurance types are balanced on those aspects regarding patients' willingness to undergo surgery. Lastly, given similar health conditions between the two types, I also show that their out-of-pocket expenses are balanced, which means their preferences related to medical treatment are close. In order to show that the preferences by insurance type are close for elective medical procedures, I restrict the sample to be the excellent-health group (self-reported) because their health care consumption may be more selective. The results are robust and unreported. On the other hand, the total medical costs are higher for the better insured, but that is exactly the moral hazard effect one would expect to arise from different coinsurance rates.

Overall, patients' health care preferences are barely associated with insurance types.

E.5 Elasticity of Surgery Probability Across Diagnoses

Table E.7: Elasticity of Surgery Probability w.r.t. Physician Financial Incentives

	Diagnoses with More Discretionary Treatment			Diagnoses with Less Discretionary Treatment			
	Neck Arthritis	Kidney Stone	Varicose Veins	Cancer	Hip Fracture	Kidney Failure	Emergency Room Visits
Elasticity of surgery probability	16.2	4.8	4.3	0.9	1.2	-0.5	-0.2

Note: *The elasticity of surgery probability is defined as the ratio of the percentage change in surgery rate over the percentage change in charge differential. The classification of services by discretion level follows that in Clemens and Gottlieb (2014).*

F Details in Hump-Shape Regression (Equation (15))

F.1 Assumptions and Coefficient Estimates for Hump-Shape Regression

The endogenous regressors include charge differential and regressors which are functions of the charge differential (higher-order terms and the out-of-pocket price). The instrument includes the policy change ($D_h \cdot Post_t$) and its interactions with coinsurance rate, age, gender, comorbidity, and subsidy status. As in the standard control function approach, I assume that in the first stage, each endogenous variable is linear in all the instruments and exogenous variables. The error terms from the first stage and the error in the main regression follow a multivariate normal distribution.

I use the ivprobit command in Stata to estimate the parameters in Equation (15) and the parameters that show the relationships between the endogenous variables and instruments in one step. This approach further assumes that the endogenous variables are continuous. Figure F.8 below indicates that it is reasonable to view the measured charge differential, $\Delta\widehat{T_{ihq}}$, as continuous.

The last assumption required is that the six instruments represent the whole set of instrumental variables for the endogenous variables. Although I cannot directly verify this assumption, I have confirmed that the estimation results are quite robust to other specifications of the set of instruments. For example, when I drop one or two instruments or add one more instrument, the interaction of the policy change and squared coinsurance rate, the coefficient changes are limited.

The estimation results are displayed in Table F.8. The bootstrap standard errors account for both the measurement of ΔT_{ihq} in Equation (14) and the two stages in the main regression (15). The Wald test of the exogeneity of the instrumented variables yields a Wald statistic of 76.45, rejecting the null hypothesis of no endogeneity.

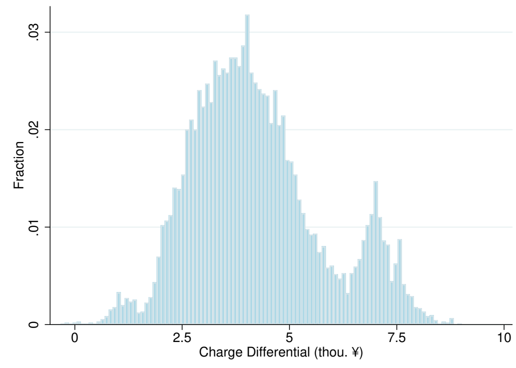


Figure F.8: Histogram for the measured charge differential ($\Delta\widehat{T_{ihq}}$), with 100 bins

F.2 Physicians' Marginal Costs Similar Between Treatment Options

I compare patients' length of stay, physicians' required time, and malpractice probability of neck arthritis between surgical and non-surgical treatment to argue that physicians' marginal costs are similar between the two treatment modes.

First, the average length of stay for surgical treatment is 14 days, versus 12 days for non-surgical treatment. Moreover, the operation time for the surgery itself is 2 to 3 hours, both in China (Wang et al., 2016) and in the U.S. (Burkhardt et al., 2013). By contrast, non-surgical treatment requires both therapy and acupuncture on a daily basis. Thus, even assuming that physicians only spend 30 minutes in total per day (there is no record of time in the data; this number is based on my interviews with physicians), the total active working time for a physician under non-surgical treatment - calculated using the average length of stay - is twice as much as that under surgical treatment. Admittedly, time spent in the operating room is more intense, but it is unclear how a well-trained professional would trade off between intensity and length of time. The bottom line here is that the overall difference in terms of physicians' time and effort seems rather limited.

As for malpractice risks, I collected the universe of malpractice litigation in the study region through China Judgments Online, an online database owned by the central government.[38] There were only 65 malpractice cases during the 4 years of observation (with more than 700,000 inpatient visits), thus indicating that the malpractice risk in general is low. More importantly, none of the lawsuits centered on neck arthritis, which leads me to deduce that the differences in terms of malpractice risks between the two treatment modes are very negligible. Since we tend to think that surgery involves higher (malpractice) risks, I leverage the malpractice RVUs

[38]The official website for China Judgments Online is: https://wenshu.court.gov.cn/.

Table F.8: Estimation Results for Specifications of Equation (15)

	(1)	(2)	(3)	% Change from (2) to (3)
Financial Incentives				
Charge Differential	0.269	0.335	0.333	-0.6%
	(0.114)	(0.167)	(0.154)	
(Charge Differential)^2	-0.034	-0.043	-0.045	-4.5%
	(0.017)	(0.021)	(0.022)	
(Charge Differential)^3	0.000	0.000	0.000	22.1%
	(0.001)	(0.003)	(0.003)	
OOP	-0.201	-0.127	-0.127	-0.5%
	(0.062)	(0.065)	(0.065)	
Patient Demographics				
Age	-0.021	-0.033		
	(0.011)	(0.016)		
Gender (male=1)	0.101	0.139		
	(0.081)	(0.266)		
Subsidy (yes=1)	-0.128	-0.196		
	(0.065)	(0.107)		
Comorbidity (yes=1)	-0.023	-0.051		
	(0.007)	(0.017)		
Patient Characteristics	Yes	Yes	No	
Hospital Fixed Effects	Yes	Yes	Yes	
Quarter-Year Fixed Effects	Yes	Yes	Yes	
Instrumental Variables	No	Yes	Yes	

Note: *Bootstrap standard errors are in parentheses, applied to the measurement of* ΔT_{ihq} *and the two stages in the main regression Equation* (15).

in the U.S. to get a sense of the risk level for spinal disk surgeries of the neck. Malpractice RVUs data is publicly available from CMS and shows that the risk level of the neck surgery - with the malpractice RVUs being 4.99 - falls somewhere between that of appendectomy and knee replacement (malpractice RVUs = 3.57 and 5.56, respectively). The surgery risk is also far lower than that for a C-section and even than the easiest Coronary artery bypass with a single arterial graft (malpractice RVUs = 11.49 and 7.95, respectively).

Overall, assuming the underlying cost difference to be zero could approximate the reality of how this diagnosis is treated in China. However, the generalizability of this assumption to other diagnoses or settings differs from case to case.

F.3 Hump-Shape Regression with Charge Differential Per Day

I illustrate that my results are robust to the alternative specification of charge differential per day. To see this, first note that if the length of stay for the two treatment options is exactly the same for every case, then the charge differential per episode is just a multiple of the charge differential per day. Correspondingly, the coefficients for charge differential per episode in equation (15) will be the same multiple of the coefficients if we use charge differential per day in the regression.

To visualize this claim, column (1) of Table F.9 displays coefficients for per-episode analysis in this regression, and column (2) of Table F.9 displays hypothetical coefficients using charge differential per day (assuming that the length of stay for the two modes are exactly the same at the current mean [12.44 days]). Basically, the hypothetical coefficient for charge differential per day is the original coefficient multiplied by the mean length of stay 12.44 ($0.335 \times 12.44 = 4.163$). Similarly, the coefficient for the charge differential per day squared is the original coefficient multiplied by the squared mean length of stay ($-0.043 \times 12.44^2 = -6.628$). In this case, per-episode and per-day analysis deliver equivalent partial correlations.

Table F.9: Coefficient Estimates for Financial Incentives Per Episode Versus Per day

	(1) Per episode	(2) Hypothetical per day	(3) Per day (spec 1)	(4) Per day (spec 2)
Financial Incentives				
Charge Differential	0.335	4.163	4.326	4.548
	(0.167)		(2.013)	(2.054)
(Charge Differential)^2	-0.043	-6.628	-7.357	-7.971
	(0.021)		(3.260)	(3.576)
(Charge Differential)^3	0.000	0.651	0.681	0.996
	(0.003)		(2.548)	(9.632)
OOP	-0.127	-1.576	-1.502	-1.549
	(0.065)		(0.781)	(0.810)
Patient Demographics				
Age	-0.033	-0.033	-0.033	-0.033
	(0.016)		(0.016)	(0.016)
Gender (male=1)	0.139	0.139	0.141	0.138
	(0.266)		(0.285)	(0.297)
Subsidy (yes=1)	-0.196	-0.196	-0.194	-0.190
	(0.107)		(0.112)	(0.116)
Comorbidity (yes=1)	-0.051	-0.051	-0.060	-0.056
	(0.017)		(0.024)	(0.021)
Hospital Fixed Effects	Yes	Yes	Yes	Yes
Quarter-Year Fixed Effects	Yes	Yes	Yes	Yes
Instrumental Variables	Yes	Yes	Yes	Yes

Note: *This table reports the coefficient estimates and their standard errors in parentheses from equation (15), using charge differential per episode and per day, respectively. The number of observations is 10,596. The bootstrap standard errors account for both the measurement of ΔT_{ihq} in Equation (14) and the two stages in the main regression (15).*

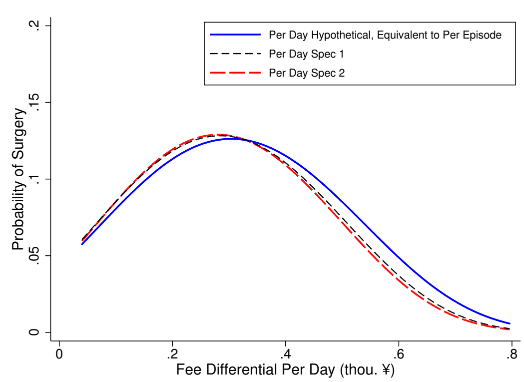
In my setting, the average length of stay is 14.25 days for surgical treatment — only 1.97 days longer than that for non-surgical treatment. The small difference in the length of stay implies that replacing charge differential per episode with per day leads roughly to a rescaling effect on the estimates. In other words, the more the coefficients under per day analysis mimic the hypothetical coefficients in column (2), the lesser the bias in using per-episode analysis. Therefore, I run the regression with charge differential per day and check whether the coefficients and predicted surgery probabilities are similar enough to the hypothetical episode-equivalent per-day specification.

I calculate the charge differential per day in two different ways. In the simplest way, I get individual daily charge for a treatment mode by dividing the current imputed total charges (from equation (14)) by the average length of stay for each treatment. For example, if $\widehat{T^s_{ihq}} = \$3,000$ and $\widehat{T^d_{ihq}} = \$1,500$, then $\widehat{T^s_{ihq}}$ per day $= \frac{\$3{,}000}{14.25}$, $\widehat{T^d_{ihq}}$ per day $= \frac{\$1{,}500}{(14.25-1.97)}$, and the daily charge differential is the difference of the two daily charges. Alternatively, I calculate the daily charge observed in the data, and impute the daily charge differential as in equation (14), using the daily charge in the data as the Y variable. In this way, individual length of stay is non-parametrically estimated.

Estimation results with per day analysis appear in columns (3) and (4) of Table F.9. Both specifications for total charges per day lead to very similar results, and the estimated coefficients for financial incentives do not deviate from the hypothetical case much, implying that the choice of per-episode analysis is inconsequential. The estimated coefficients for patient demographics are almost the same across different specifications.

I then use the coefficients to graph the relationship between surgery rates and charge differential per day in Figure F.9 for the two per-day specifications and the hypothetical per-day analysis, which ignores any differences in the length of stay between the two modes as in the per-episode analysis. All the specifications lead to a similar pattern. Intuitively, assuming the length of stay for the two modes to be the same will slightly overstate the charge differential per day, pushing the surgery distribution under hypothetical per-day analysis slightly to the right of the other two specifications. However, it only shifts to a limited extent, given the similar length of stay between the two modes.

Figure F.9: Surgery Probability and Charge Differential Per Day



F.4 Regression Using Readmission Indicator

Based on Property 2, conditional on a patient who opts for surgery, the expected value of ξ_i should decrease and then increase with ΔT_i. Ideally, this pattern can be tested using a proxy for ξ_i, for example, pain level or clinical measurement for the suitability of surgery. The idea of a proxy for ξ_i is similar to Chan Jr et al. (2019), where they use a radiologist's share of patients who leave with undiagnosed pneumonia as the proxy for the accuracy of the radiologist's received signal.

The only health outcome I have is the readmission rate. However, it does not qualify for an ideal proxy from either a clinical or statistical point of view, given its average value of 1 percent. Nonetheless, I plot the 30-day readmission rate against the charge differential by replacing the surgery dummy with the readmission dummy in equation (15) and using the sub-sample who chose surgery. As shown in Figure F.10, the hump shape indeed exists, although the estimates are not precise given this rare event.

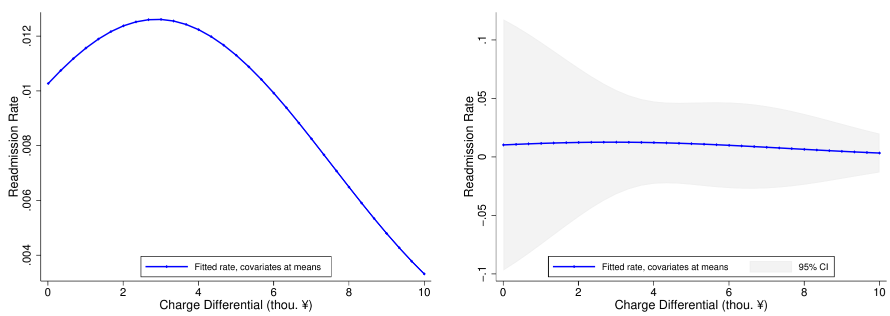


Figure F.10: Readmission Rate for Surgery Patients and Conflict of Interest

Note: *Conflict of interest is measured by the charge differential. The fitted relationship is conditional on patient demographics, time trends, and hospital quality. The figure on the left drops the confidence interval so that the hump shape can be visualized. Confidence intervals are based on bootstrap, accounting for the measurement of ΔT_{ihq} in Equation (14) and the two stages in the main regression.*

G Incorporating Health Outcomes in the Bayesian Persuasion Model

Patient i's health outcome, R_i^*, can be different between surgery (s) and drugs (d). Specifically,

$$\begin{aligned} R_i^* &= \rho_0 + \boldsymbol{\rho}_1 \cdot \boldsymbol{Z_{1i}} + r_i \cdot \mathbb{1}(s_i = 1) + v_i, \\ \text{where } r_i &= r_0 + \boldsymbol{r}_1 \cdot \boldsymbol{Z_{1i}} + \xi_i, \end{aligned} \tag{G.1}$$

where R_i^* is a continuous measure for health outcome, which could represent the level of pain. r_i captures the relative health benefit of surgery, which is assumed to be a function of health-related demographics $\boldsymbol{Z_{1i}}$, such as age and indication for comorbidity, and ξ_i, the random coefficient for choosing surgery. In other words, the error term ξ_i from my original model now represents the potential health outcome between the treatment options. The error term v_i captures any unobserved health factors.

I now modify the patient utility from surgery (utility from drugs is normalized to be 0). The relative value of s for patient i contains two parts. The first part is the relative health benefit, measured by the random coefficient r_i. The second part is other preferences for surgery, denoted as $\tilde{\kappa}_i$, which is a function of i's characteristics. The assumption here is that the health component and preference component are separately additive.

$$\begin{aligned} U_i^s &= r_i - \tilde{\kappa}_i \\ &= r_0 + \boldsymbol{r}_1 \cdot \boldsymbol{Z_{1i}} + \xi_i - \tilde{\kappa}_i. \end{aligned} \tag{G.2}$$

Since patients' health-related demographics, $\boldsymbol{Z_{1i}}$, are a subset of i's characteristics, I define $\kappa_i \equiv -(r_0 + \boldsymbol{r}_1 \cdot \boldsymbol{Z_{1i}}) + \tilde{\kappa}_i$, to capture all the utility from surgery that is related to patient characteristics. As a result,

$$U_i^s = \xi_i - \kappa_i. \tag{G.3}$$

Although the utility form looks the same as before, ξ_i now has a clear interpretation of medical benefit. Its distribution can be estimated using observed health outcomes jointly with other model parameters with MLE. The joint framework is such that:

In the first stage, patient i chooses surgery ($s_i = 1$) if and only if $U_i^s > U_i^d$; that is, $\xi_i \geq c_i^*$ where c_i^* is the optimal cutoff derived from the persuasion model.

In the second stage, patient i is fully recovered ($R_i = 1$ or readmission indicator = 0) if and only if $R_i^* > 0$; that is, $v_i > -\rho_0 - \boldsymbol{\rho}_1 \cdot \boldsymbol{Z_{1i}} - (r_0 + \boldsymbol{r}_1 \cdot \boldsymbol{Z_{1i}} + \xi_i) \cdot \mathbb{1}(s_i = 1)$.

Note that the second stage has assumed that the potential health outcome producing the recovery dummy is exactly the relative health benefit in patient utility, primarily because readmission status is the only health outcome measure in my data (and in many other comparable datasets). However, one might view ξ_i as other medical outcomes instead, such as the level of pain during and/or after the treatment.

Regarding estimation, in practice, some parametric assumptions about the distribution of the error terms ξ_i and v_i would be helpful for the tractability of MLE. For example, if ξ_i and v_i are bivariate normal,

$$\begin{bmatrix} v_i \\ \xi_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ \mu_\xi \end{bmatrix}, \begin{bmatrix} 1^2 & \rho\sigma_\xi \\ \rho\sigma_\xi & \sigma_\xi^2 \end{bmatrix}\right),$$

then the individual likelihood $Pr(s_i = d^s, R_i = d^r|\cdot)$, where $d^s, d^r \in \{0, 1\}$, has an analytical form.

H Analytical Expressions for Treatment Outcomes and Expected Welfare

H.1 Treatment Outcomes

For patient i, plugging the estimated coefficients into equation (20), the model's predicted optimal cutoff is

$$\begin{aligned} \hat{c_i^*}(\Delta\widehat{T_{ihq}}, y_i, \hat{\tau}_h, \hat{\psi}_q, \mathbf{X_i}, \widehat{\nu_{ihq}}) = &\max\Big\{ -\lambda^{-1}\big(\hat{\alpha_1} \cdot y_i \cdot \Delta\widehat{T_{ihq}} + \hat{\alpha_2} \cdot (y_i \cdot \Delta\widehat{T_{ihq}})^2 + \hat{\tau}_h + \hat{\psi}_q + \hat{\boldsymbol{\gamma}} \cdot \mathbf{X}_i + \hat{\zeta}_1 \cdot \widehat{\nu_{ihq}}\big), \\ &(\hat{\alpha_1} \cdot y_i - \hat{\beta}_1) \cdot \Delta\widehat{T_{ihq}} + (\hat{\alpha_2} \cdot y_i^2 - \hat{\beta}_2) \cdot \Delta\widehat{T_{ihq}}^2 + \hat{\tau}_h + \hat{\psi}_q + \hat{\boldsymbol{\gamma}} \cdot \mathbf{X}_i + \hat{\zeta}_2 \cdot \widehat{\nu_{ihq}}\Big\}. \end{aligned} \tag{H.1}$$

The predicted probability of surgical treatment and the expected total expenditure for i are

$$\begin{aligned} &Pr(sur_i = 1) = Pr(\xi_i \geq \hat{c_i^*}) = 1 - \Phi(\hat{c_i^*}), \\ &\widehat{T_{ihq}} = Pr(sur_i = 1) \cdot \widehat{T^s_{ihq}} + \big(1 - Pr(sur_i = 1)\big) \cdot \widehat{T^d_{ihq}}, \end{aligned} \tag{H.2}$$

where $\widehat{T^s_{ihq}}$ and $\widehat{T^d_{ihq}}$ are obtained from regression (14) by turning the surgery indicator, $\mathbb{1}(\boldsymbol{s_i = 1})$, on and off.

The average likelihood of surgery and average total expenditure among N patients are

$$\begin{aligned} &Pr(sur = 1) = \frac{\sum_{i=1}^N \big(1 - \Phi(\hat{c_i^*})\big)}{N}, \\ &\hat{T} = \frac{\sum_{i=1}^N \widehat{T_{ihq}}}{N}. \end{aligned} \tag{H.3}$$

H.2 Monetize Patient Utility

To get welfare in monetary terms, patient utility is normalized by the marginal utility of money. Since the utility of receiving z dollars is $\hat{\alpha_1} \cdot z - \hat{\alpha_2} \cdot z^2$, the marginal utility of money is $\hat{\alpha_1} - 2\hat{\alpha_2} \cdot z$, which varies with the wealth level. For each patient, I use their out-of-pocket price as the reference wealth level. That is, patient i's marginal utility of money is specified as

$\hat{\alpha_1} - 2\hat{\alpha_2} \cdot y_i \cdot \widehat{\Delta T_{ihq}}$. Replacing the individual out-of-pocket price $y_i \cdot \widehat{\Delta T_{ihq}}$ with the average out-of-pocket price has been shown to change the results rather modestly.

The welfare of non-surgical treatment is normalized to be 0. Therefore, all the welfare levels are relative to those of non-surgical treatment. Given the model estimates, patient i's expected welfare is

$$\begin{aligned}\mathbb{E}\big[\widehat{W_i^{patient}}\big] =& Pr(\xi_i \geq \hat{c_i^*}) \times \frac{\mathbb{E}(\xi_i|\xi_i \geq \hat{c_i^*}) - \hat{\kappa_i}}{\hat{\alpha_1} - 2\hat{\alpha_2} \cdot y_i \cdot \widehat{\Delta T_{ihq}}} + Pr(\xi_i < \hat{c_i^*}) \times 0 \\ =& (1 - \Phi(\hat{c_i^*})) \times \frac{\frac{\phi(\hat{c_i^*})}{1-\Phi(\hat{c_i^*})} - \hat{\kappa_i}}{\hat{\alpha_1} - 2\hat{\alpha_2} \cdot y_i \cdot \widehat{\Delta T_{ihq}}}\end{aligned}, \tag{H.4}$$

where $\hat{\kappa_i} = \hat{\alpha_1} \cdot y_i \cdot \widehat{\Delta T_{ihq}} + \hat{\alpha_2} \cdot (y_i \cdot \widehat{\Delta T_{ihq}})^2 + \hat{\tau_h} + \hat{\psi_q} + \hat{\boldsymbol{\gamma}} \cdot \mathbf{X}_i + \hat{\zeta_1} \cdot \widehat{\nu_{ihq}}$.

The total expected welfare of the patients in the sample is

$$\mathbb{E}\big[\widehat{W^{patient}}\big] = \sum_{i=1}^{N} \mathbb{E}\big[\widehat{W_i^{patient}}\big]. \tag{H.5}$$

I Comparison to a Model Without Obedience Constraint

I.1 The Alternative Model

While the patient in the persuasion model plays an active role by rationally judging the recommendation and has the right to disagree (through the obedience constraint), in a model without the obedience constraint, the patient's skepticism is at most implicit through the weight on the patient's utility. Therefore, one can view the alternative model as where the obedience constraint in Equation (8) never binds. In this case, $c_i^* = \kappa_i - F_i$. Intuitively, without the obedience constraint, the physician's optimal strategy is to choose surgery if and only if surgery gives weakly higher utility, i.e., $\xi_i - \kappa_i + F_i \geq 0$. In other words, the optimal cutoff $c_i^* = \kappa_i - F_i$.

I.2 Different Model Properties

I contrast the four model properties between the two models.

For model Property 1, which addresses how physicians' response to reimbursement depends on patient insurance, in the alternative model, the cross partial $\frac{\partial}{\partial \kappa_i}\left(\frac{\partial c_i^*}{\partial F_i}\right) = 0$. This means that in the alternative model, physicians' response to financial incentives is independent of insurance generosity.

For model Property 2, which discusses the hump-shaped relationship between surgery probability and the charge differential ΔT_i, this relationship is not a robust prediction in the alternative model.

In the alternative model, the derivative of cutoff w.r.t. charge differential is

$$\frac{\partial(c_i^*)}{\partial(\Delta T_i)} = \frac{\partial k_i}{\partial \Delta T_i} - f'(\Delta T_i). \tag{I.1}$$

In order to generate the non-monotonic relationship between surgery probability and charge differential in Property 2, the alternative model needs to assume some preferences such that the derivative in (I.1) is negative below some level of charge differential, and then it becomes positive. The chosen preferences may be justified by the patient's or the physician's changing marginal utility of financial incentives, such as patient or physician income effects. However, Section I.4 suggests that when taking this model to data, the implied signs and magnitudes of such non-linearity in the preferences could be inconsistent with what theory and empirical studies indicate.

Regarding Property 3 on the decomposition of moral hazard and greater misdirection effects, since the ability to decompose the effects of improved cost-sharing into moral hazard and greater misdirection arises only when the obedience constraint binds, Property 3 does not exist in the alternative model. In this model, the misdirection is $k_i - (k_i - F_i) = F_i$, which is independent of patient cost-sharing. This lack of interaction between cost-sharing and misdirection in the alternative model contrasts with the Bayesian persuasion framework, where the misdirection effect is directly influenced by patient insurance generosity.

Regarding Property 4 on the equivalence between shutting down physician financial incentives and eliminating information asymmetry, the Bayesian model suggests that either mitigating information asymmetry, which increases patients' skepticism, or reducing the relative importance of physicians' financial sensitivity can limit physicians' ability to push for unnecessary surgery. In contrast, in the alternative model, physicians' ability to distort patients' decisions can only be limited by decreasing the weight of physician financial incentives. The first best allocation from the patient's perspective can only be achieved by shutting down physicians' incentives.

I.3 Estimating a Model without Obedience Constraint

To gauge the importance of patients' explicitly active roles in the Bayesian persuasion model, I estimate the model without the obedience constraint, with the same utility specifications and control function. That is, I estimate the parameters governing physicians' weights on patients, and assume a quadratic form of patient out-of-pocket price and physician's sensitivity for money to capture potential patient and physician income effects. For robustness check, I also relax the quadratic specification to higher-order polynomials to allow for more flexibility and to rule out the mechanical effects of the functional form. Based on the main specification, the optimal cutoff for surgery in the alternative model is

$$c_i^*(\widehat{\Delta T_{ihq}}, y_i, \tau_h, \psi_q, \mathbf{X_i}, \widehat{\nu_{ihq}}) = (\alpha_1 \cdot y_i - \beta_1) \cdot \widehat{\Delta T_{ihq}} + (\alpha_2 \cdot y_i^2 - \beta_2) \cdot \widehat{\Delta T_{ihq}}^2 + \tau_h + \psi_q + \boldsymbol{\gamma} \cdot \mathbf{X}_i + \zeta_2 \cdot \widehat{\nu_{ihq}}. \quad \text{(I.2)}$$

This is exactly the expression in the second component of the max function, which specifies the optimal cutoff in the Bayesian persuasion model (equation (20)). In other words, if the parameters are such that the obedience constraint is slack for everyone, the Bayesian persuasion model is equivalent to a model where a partially altruistic physician makes the treatment decision alone. Therefore, a likelihood ratio test can be used to assess the goodness of fit of the two competing models.

The estimates are displayed as follows in Table I.10.

Table I.10: Structural Model Parameter Estimates

	Bayesian Persuasion		Without Obedience Constraint	
	(1)	(2)	(3)	(4)
Financial Incentives				
β_1 : Charge Differential	0.004	0.154	0.390	0.527
	(0.001)	(0.065)	(0.143)	(0.194)
β_2 : (Charge Differential)^2	-0.002	-0.001	-0.045	-0.057
	(0.001)	(0.001)	(0.026)	(0.034)
α_1 : OOP	0.080	0.065	0.479	0.392
	(0.028)	(0.033)	(0.135)	(0.149)
α_2 : OOP^2	0.004	0.003	-0.020	-0.011
	(0.003)	(0.005)	(0.008)	(0.006)
Patient Demographics				
γ_1 : Age	0.027	0.041	0.032	0.036
	(0.008)	(0.015)	(0.014)	(0.018)
γ_2 : Gender (male=1)	-0.139	-0.084	-0.130	-0.144
	(0.083)	(0.058)	(0.072)	(0.146)
γ_3 : Subsidy (yes=1)	0.019	0.074	0.106	0.190
	(0.009)	(0.062)	(0.048)	(0.100)
γ_4 : Comorbidity (yes=1)	0.002	0.014	0.046	0.063
	(0.001)	(0.006)	(0.013)	(0.031)
ζ_1 : Residual for $\widehat{\Delta T_{ihq}}$ in the First Argument of Max Function		0.106		
		(0.052)		
ζ_2 : Residual for $\widehat{\Delta T_{ihq}}$ in the Second Argument of Max Function		0.022		0.002
		(0.012)		(0.001)
Hospital Fixed Effects	Yes	Yes	Yes	Yes
Quarter-Year Fixed Effects	Yes	Yes	Yes	Yes
Instrumental Variables	No	Yes	No	Yes

Note: *Bootstrap standard errors in parentheses, applied to the measurement of* ΔT_{ihq} *in equation (14) and both stages of the maximum likelihood estimation.*

I.4 Comparison of Models

The key value-added of the Bayesian persuasion model is to inform a new trade-off in designing insurance policy, i.e., a decomposition of the downside of insurance into patient price sensitivity (the traditional moral hazard) and greater misdirection. One can only learn this decomposition by explicitly modeling patient choice. In contrast, in a model without the obedience constraint, regardless of utility specifications, the misdirection is independent of patient insurance coverage. Therefore, later in this section, I show that the two models have different welfare implications when considering changing patient coinsurance rates.

Before going into the details of this decomposition and welfare comparisons, which are robust to parametric assumptions, I examine which model can be better supported by the data and yield a more sensible counterfactual analysis under reasonable parametric restrictions.

Data Favors the Bayesian Persuasion Model Under Reasonable Parametric Restrictions

While the Bayesian persuasion model will replicate the hump-shape relationship between the surgery probability and charge differential robustly through the mechanism of obedience constraint, the alternative model can only generate the hump shape by adding non-linearity, for example, polynomials of the financial incentives. Even so, the estimates in Table I.10 indicate that the patients face lower marginal disutility of a higher out-of-pocket price in the alternative model, which is inconsistent with empirical studies documenting patients' income effects (Chernew et al., 2008; Chen and Lakdawalla, 2019).

Given the testable restriction discussed in Section I.3, the likelihood ratio test rejects the null and implies that the obedience constraint is empirically relevant in my main parameterization. Indeed, 73 percent of the physicians' constraints are binding when recommending surgery. Under the assumption that patients always follow recommendations, removing obedience constraint in the Bayesian persuasion model raises surgery probability from 8 percent to 20 percent.

Intuitively, the better fit of the Bayesian persuasion model comes from its extra disciplinary force on physicians through the obedience constraint. Therefore, when the charge differential gets higher, the marginal utilities change (incentive mechanism), and the constraint becomes binding (information mechanism), both forcing the probability of surgery to decrease. In contrast, to generate the hump-shape relationship between surgery probability and charge differential, the alternative model has to heavily rely on the higher order terms in the utility function, implying unrealistic patient and physician income effects.

Note that if the goal is just to match the data, adding flexibility to the alternative model will do as it eventually implies a non-parametric function. For example, in alternative specifications, I replace the quadratic specification in both out-of-pocket price and charge differential with higher degree polynomials to allow for even more flexibility. There is still a significant difference in the estimated likelihood between the two models, using up to fourth-order polynomials. For even higher-order terms, the obedience constraint becomes statistically trivial. The failure to reject the null hypothesis in the likelihood ratio test could simply be attributable to insufficient statistical power with the fifth-order polynomial. Even though the two models become indistinguishable under more flexible preference specifications, the Bayesian persuasion model has the advantage of parsimony and helps to avoid ad-hoc preference specifications.

In the following two comparisons, I argue that the Bayesian persuasion model appears important not because it matches data better, but because it captures policy-relevant effects that are overlooked by models without the explicit role of patient choice. These arguments are in the spirit of Berry (1992), which compares a structural model that incorporates strategic interactions among firms with simpler models in studying entry in the airline industry.

Counterfactual Analysis Regarding Physicians' Remuneration

I compare the two models' predictions on utilization and patient welfare for the counterfactual policy in Section 6.3, when physicians can get all the hospital revenue they create.

Table I.11 presents counterfactual outcomes using the main specification of a quadratic form for the financial incentives. The alternative model predicts a surgery rate decreasing to 1.4 percent (from 8.2 percent at baseline), an 82 percent drop. In contrast, the Bayesian persuasion model predicts that the surgery rate will increase to 10.1 percent (from 8.3 percent at baseline), a 22 percent rise. In reality, physicians in treated departments responded by performing 21 percent more surgeries Gong et al. (2021). Although the real policy's effect is not specific to neck arthritis, the predictions of the Bayesian persuasion model seem much more reasonable, while both the sign and magnitude are off in the alternative model's prediction due to unrealistically large physician income effects: the physician's utility from income diminishes so rapidly that they start to receive negative utility from money on the right tail of the observed financial incentives. The contrast in care utilization leads to opposite implications in patient welfare; while the patients lose due to greater distortion when incentives are more misaligned in the Bayesian persuasion model, patients are overall better off in the alternative model despite the fact that some are now misdirected toward drugs.

Notably, the predictions regarding this policy remain almost the same using the polynomial of degree five for both models. With this flexible specification, the Bayesian persuasion model predicts a 25 percent rise in surgery probability, while the alternative model predicts an 81 percent drop.

Table I.11: Comparing Counterfactual Predictions to Reality

	Baseline Model		Predictions for Policy Change in 2016		Real Change in 2016
	Without Constraint	BP	Without Constraint	BP	
Surgery Probability (%)	8.2	8.3	1.4	10.1	
Percent Change Relative to Baseline			-82	22	21
Total Expenditure (Mill. $)	15.5	15.8	14.5	16.0	
Percent Change Relative to Baseline			-7	1	3
Patient Utility	-390.2	162.2	-83.1	31.7	
Percent Change Relative to Baseline			79	-80	-

Note: *"BP" stands for the Bayesian persuasion model. Model predictions are based on my data for neck arthritis. Numbers for real change are from Gong et al. (2021). Real total expenditure change is calculated based on reported changes in daily expenditure and length of stay.*

Decomposing Moral Hazard and Greater Misdirection Effect

The counterfactual policy is to reduce every patient's coinsurance rate by half. In the misdirection-only scenario, the movement of the cutoff is $|(h^{-1}(\kappa_i^{'}) - \kappa_i^{'}) - (h^{-1}(\kappa_i) - \kappa_i)|$ in the Bayesian persuasion model with constrained physicians, and $|(\kappa_i^{'} - F_i - \kappa_i^{'}) - (\kappa_i - F_i - \kappa_i)| = 0$ in the alternative model.

Figure I.11 presents the percent changes in outcomes for the two models. In the Bayesian persuasion model, both moral hazard and greater misdirection based on lower coinsurance rate explain the rise of surgeries, with the misdirection effect accounting for about one-fifth of the increase. In the alternative model, moral hazard is the sole driving force. Perhaps the most striking difference is how the models decompose patients' out-of-pocket expenditure and welfare gains. Although patients are generally better off due to lower out-of-pocket prices, physicians' greater misdirection in the Bayesian persuasion model means that some patients waste out-of-pocket money and lose utility due to surgeries they would not have opted for otherwise, nearly offsetting the welfare gain. However, the alternative model ignores this policy-relevant effect and estimates a much higher patient welfare gain.

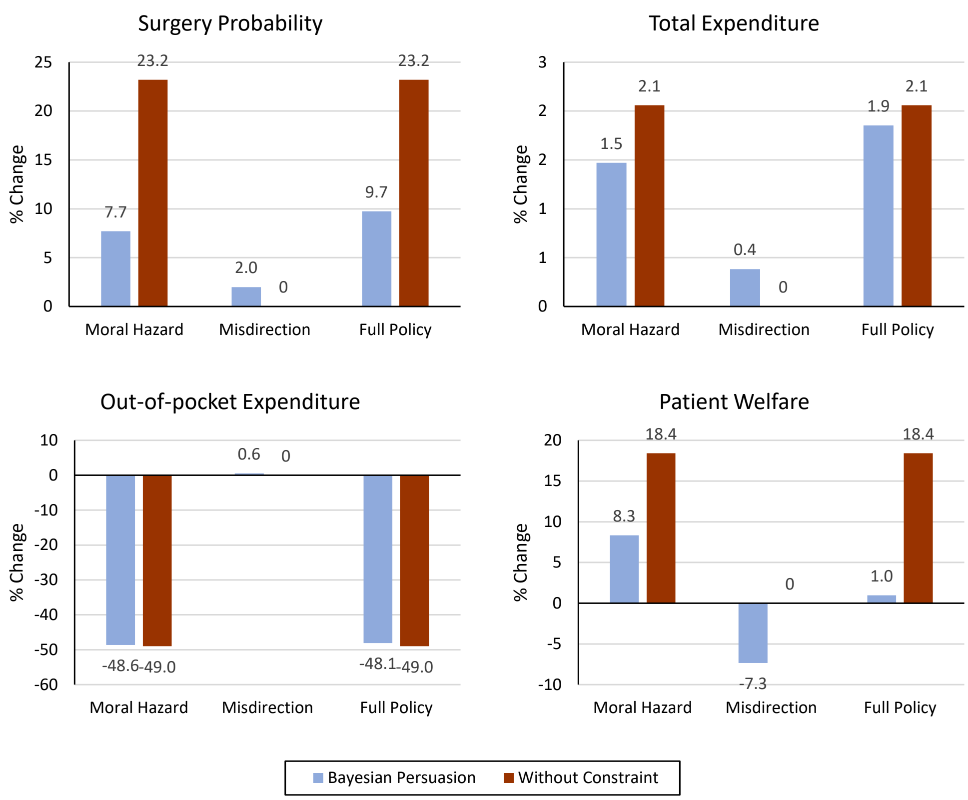


Figure I.11: Decomposition When Coinsurance Rate Decreases by Half

Note: *The figure shows the percent change in outcomes under the moral hazard-only, misdirection-only, and full-policy scenarios, respectively. The baseline case is when everyone's coinsurance rate is as it is in the data. The counterfactual case is when they face half of these coinsurance rates. Both models use a quadratic parameterization of patient out-of-pocket price and charge differential.*