Received: 9 June 2020 | Revised: 25 June 2021 | Accepted: 14 July 2021

DOI: 10.1002/hec.4407

RESEARCH ARTICLE

Health Economics WILEY

Patients' free choice of physicians is not always good

Xinyu Li[1] | Christian Waibel[2]

[1]Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands

[2]Department of Management, Technology, and Economics, ETH Zurich, Zurich, Switzerland

Correspondence
Christian Waibel, Department of Management, Technology, and Economics, ETH Zurich, Zurichbergstrasse 18, 8092 Zurich, Switzerland.
Email: cwaibel@ethz.ch

Funding information
Open Access Funding provided by Eidgenossische Technische Hochschule Zurich

Abstract

We present a model of learning in healthcare markets. Hospitals have junior physicians with low and senior physicians with high ability. Junior physicians turn senior if they treat enough patients. Patients face heterogeneous costs for waiting if a physician's capacity is utilized. Hospitals choose to either allocate patients to physicians randomly or let patients choose their physicians. In a *monopolistic* market, the hospital always chooses the welfare-maximizing allocation system. In a *competitive* market, inefficiencies may arise due to two externalities. If patients are free to choose their physician, the marginal patient neither internalizes her impact on other patients' waiting costs nor the learning of junior physicians.

KEYWORDS
healthcare markets, learning, quality, regulation, social welfare

1 | INTRODUCTION

Improving the quality of healthcare is a key objective in the design of healthcare markets. Empirical evidence suggests that physicians' experience is an important element in ensuring the quality of medical treatment. Junior physicians provide a lower quality of treatment than senior physicians. Mortality and morbidity rates are significantly higher in cohort turnover months—often referred to as the "August Killing Season"—than in other months (Young et al., 2011). Junior physicians can improve their skills and reach a senior level by treating a sufficiently large number of patients. However, patients prefer treatment by a senior physician.[1] Yet, rejecting a junior physician inhibits these physicians' learning process. Thus, the rejection of junior physicians reduces welfare in the long term. Our paper investigates the conditions that lead to reduced social welfare when patients have the freedom to choose their physician.

We employ a theoretical model to compare social welfare outcomes in a monopolistic market with those in a competitive market. There is one hospital with two junior and two senior physicians in the monopolistic market. There are two symmetric hospitals in a competitive market, each with one junior and one senior physician. Senior physicians provide a higher quality of treatment than junior physicians. If a junior physician treats enough patients during the first period, he will be a senior physician during the second period. Each physician has the same capacity to treat a quarter of the patient population. If the demand for a physician exceeds his immediate capacity, patients visiting this physician incur heterogeneous waiting costs. Waiting costs increase in parallel with excess demand. Each hospital either lets patients freely choose their physician or matches patients randomly to physicians. We refer to the former as a *free-choice system* and the latter as a *random-allocation system*.

Our model reflects a public health system in which treatment prices are regulated, and price discrimination between junior and senior physicians is impossible. Hospitals announce their allocation system at the beginning of each of the two periods. Patients then decide which hospital to visit and, in the free-choice system, which physician to consult. Both the random-allocation and the free-choice systems are used in European countries. In Germany, Austria, and Switzerland, the two systems co-exist. Privately insured patients can freely choose a hospital and a physician to visit. In contrast, publicly insured patients can freely choose a hospital, but they are randomly assigned to one of the hospitals' physicians.

Our paper contributes to the design of healthcare markets. In particular, we add to the existing literature by providing a model that allows us to identify the welfare-maximizing allocation system for any distribution of junior and senior physicians. Our results show that the hospital is indifferent to whether it implements the free-choice system or the random-allocation system in a monopolistic market. The reason is that under both systems, all patients are treated by the same hospital. Under a free-choice system, more patients are treated by a senior physician than a junior physician. Only under the random-allocation system do junior physicians treat enough patients in the first period to become senior in the second period. Hence, the policy-maker trades off between offering patients high-quality treatments in the first period and enhancing the learning of junior physicians for the second period.

We show that under competition, hospitals implement the system that maximizes social welfare if the quality difference is either large or small. In the case of a large quality difference, both hospitals and policy-makers prefer the free-choice system. In the case of a small quality difference, hospitals and policy-makers prefer to implement the random-allocation system. However, for intermediate quality differences, hospitals prefer the free-choice system over the random-allocation system although the random-allocation system would lead to a higher level of social welfare. The deviation of market equilibrium from the socially optimal choice stems from market imperfection.

Patients' choices may generate a positive or negative externality for other patients. Under a free-choice system, the marginal patient whose visit makes a junior physician senior generates a positive externality for all other patients in the next period. However, when patients are waiting for a physician, each additional visiting patient imposes a negative externality on other waiting patients. The competition between hospitals creates a Prisoner's Dilemma in which choosing a free-choice system becomes the dominant strategy although the socially optimal outcome would be that both hospitals choose a random-allocation system.

From a more general perspective, healthcare quality can be measured by multiple dimensions. Patients' allocation to physicians is just one dimension of quality in healthcare markets. Other dimensions of quality include the hospital's interior, the quality perceived by patients, the waiting time, and the actual quality of care, for instance. If patients shift focus from one dimension of quality to another, the market outcomes will be different (see Huesmann and Mimra [2021]). To the best of our knowledge, we are the first to study the impact of the allocation systems in hospitals on social welfare. A remarkable result in our model is that competition in the healthcare market does not necessarily improve care quality, which differs from previous literature. In the previous literature, when patients can freely choose their treatment providers, competition among providers triggers quality improvement because the price is regulated, and quality is the only factor that providers can use to attract more patients. If physicians can improve the quality of care without treating patients directly, for example, by theoretical knowledge, each physician can improve his care quality no matter how many patients he treats. However, if the quality of care can be improved by physicians exclusively treating patients, competition does not always lead to quality improvement. Hospitals trade-off the practical experience of their physicians and their hospitals' quality of care. If patients can freely choose their physician, they prefer to visit senior over junior physicians due to these physicians' higher ability. Nevertheless, senior physicians who cannot improve their skills anymore serve more patients. Hence, junior physicians who could improve their skills lack chances to improve their quality of care. This generates a loss of welfare.

This paper relates to at least three strands of the literature: the literature on patients' free choice and quality in healthcare markets, learning, and public policy.

Quality: The paper most closely related to our paper is Brekke et al. (2008). The authors investigate the impact of patients' free choice of hospitals (competition) on waiting times in a healthcare market with regulated prices. Brekke et al. make use of the static Salop model with heterogeneous patients. Then the introduction of competition may have an ambiguous effect on waiting time. Only when the competitive demand segment is sufficiently small, hospital competition reduces waiting times. In a follow-up article, Sá et al. (2019) allows for dynamics in the above environment. Additionally, the authors assume that a policy-maker can punish hospitals for long waiting times. They show that as long as hospitals can be held liable for waiting times, increasing patients' choices also leads to increases in hospitals' waiting times in equilibrium. Greater patient choice makes demand more responsive to changes in waiting times. The authors show that a unilateral reduction in waiting time at a hospital leads to a larger demand for that hospital.

Beitia (2003) investigates patients' hospital choice in a market with regulated prices. The benevolent regulator may not be able to enforce the hospitals' qualities. Furthermore, the regulator may have incomplete information. The author shows that competition between hospitals via quality can be used as a device to maintain an appropriate level of services in a duopoly market. Hehenkamp and Kaarbèe (2020) show that whether the competition from the private for-profit hospitals increases the public hospital's quality depends on the regulated price level.

Ma (1994) focuses on the trade-off between quality provision and cost retainment in health care. Ma shows that a first-best prospective payment exists if and only if physicians cannot refuse the treatment of high-risk patients. Chalkley and Malcomson (1998) show that physicians' payments may differ between capitation and cost reimbursement. They show that the optimal contract depends on providers' preferences. If the provider is entirely benevolent, incentives between physicians and patients are aligned. Consequently, an entirely benevolent provider offers optimal patient care. If the provider is entirely selfish, the opposite is true: preferences are aligned against each other. Therefore, total cost-sharing would be the optimal payment scheme for entirely selfish physicians.

Brekke et al. (2007) study the impact of gate-keeping on quality in healthcare markets. General practitioners act as gatekeepers. Gatekeepers separate between minorly ill patients and heavily ill patients. The authors show that compulsory gate-keeping may lead to excessive quality competition in the secondary market. Brekke et al. (2011) show that in a model of price regulation, introducing price competition among hospitals may have ambiguous effects on quality provision.

The empirical literature mainly supports the finding that patients choose their hospital based on the hospitals' quality (Howard, 2006; Tay, 2003; Varkevisser et al., 2012). Cooper et al. (2011) exploits a reform by a natural field experiment in Great Britain to show that competition among hospitals increases quality, measured as a reduction in mortality. Brekke et al. (2014) provide an excellent overview on patients' choice of hospitals.

Learning: Sobel (2000) defines learning as making a suboptimal decision in one period in order to obtain information that will improve future decision-making. Gong (2017) shows that this learning-by-doing plays a crucial role in skill acquisition by junior physicians. Our paper uses the learning-by-doing concept. Whereas Gong (2017) aims to differentiate different channels of learning, our analysis will reveal whether a learning process is welfare optimal, given the difference in physicians' ability. Narayanan and Manchanda (2009) point out that junior physicians might not necessarily learn homogeneously. For simplicity, however, we will focus on a model in which the learning experience of physicians is identical.

Physicians' learning could be driven by extrinsic rewards or intrinsic motives (Prendergast, 1999). As Rebitzer and Taylor (2011) argue, the incentives provided by employers are essential for employees to improve their skills. In our model, instead of investigating the motives of learning, we focus on how the learning process is influenced by the choice of patient allocation system.

Public policy: As to public policy in healthcare markets, Cutler (1996) argues that there are two fields in which governments may want to intervene: externalities arising from individuals' behavior and distortions in markets for medical care. Our paper shows that such externalities may arise in a free-choice system if physicians are capacity constrained. Cuff et al. (2012) compare individuals' willingness to pay for budget-constrained public health insurance under a need-based allocation versus a random allocation of budget resources. In a follow-up article, the authors test their theoretical model (Buckley et al., 2012). In contrast to Cuff et al. (2012) and Buckley et al. (2012), our focus is on the allocation of patients to physicians rather than the rationing of access to care. We allow all patients to be treated, but waiting times may delay treatment.

Section 2 presents a basic model with a monopolistic hospital. Section 3 introduces competition to the healthcare market. Section 4 describes the subgame perfect Nash equilibrium and its implications for social welfare. Section 5 discusses our model assumptions and possible extensions. Section 6 concludes the study.

2 | BASIC MODEL: HOSPITAL ACTS AS MONOPOLIST

Players: There is a payoff-maximizing monopolistic hospital with four physicians. The physicians are heterogeneous in experience. Two of them are junior physicians and two are senior. Senior physicians always cure patients. The junior physicians cure patients with probability $0 < a < 1$. Learning-by-doing allows junior physicians to become senior physicians if they treat enough patients. We assume that junior physicians have to treat at least a quarter of the population to attain senior status.[2] Each physician is capacity constrained and can only serve a quarter of the population. If the number of patients exceeds the physician's immediate capacity, patients have to wait. Ex-ante, it is not apparent whether senior physicians have higher costs in treating patients than junior physicians. On the one hand, costs might be higher for senior physicians because their time is more costly; on the other hand, senior physicians may have lower costs because their diagnosis and treatment are much quicker. Hence, we stick to the identical costs for junior and senior physicians. However, the qualitative results are robust to heterogeneity in costs. Furthermore, we normalize the physicians' costs for providing treatment to zero.

There is a continuum of payoff-maximizing patients with mass 1. Patients observe quality differences of physicians, which is supported by empirical evidence (see literature review in Section 1). Patients are heterogeneous with respect to the costs of waiting if that physician's capacities are exhausted. A patient's waiting cost is given by θT, where T is the average waiting time, and θ is the degree of the patient's dislike to wait. θ follows a uniform distribution over [0,1]. The dislike-to-wait index θ is information private to each patient, so that a hospital cannot sort patients based on θ. We assume that health insurance is mandatory for all patients. Furthermore, the health insurance company fully covers all healthcare expenses.

Actions: The hospital chooses to implement either the *random-allocation system* or the *free-choice system*. Implementing the random-allocation system (i) allows junior physicians to learn and (ii) reduces the likelihood that patients will have to wait for treatment. Implementing the free-choice system increases the patients' demand for treatment as long as the senior physicians have not yet exhausted their capacities. Patients choose their physician if the hospital implements a free-choice system.

Payoffs: The hospital's payoff π is determined by the mass of patients who visit the hospital. The hospital receives a payment p from the health insurance company for each treatment provided by a physician,[3] no matter whether the patient is cured or not. As fees are regulated, we assume that price discrimination between junior and senior physicians is impossible. We provide supporting evidence for this assumption as well as a discussion in Section 5. The patients' payoff depends on whether treatment was successful. Successful treatment leads to a value of 1 for the patient; otherwise, the treatment has a value of 0. If a patient does not visit a physician, she obtains an outside value of 0. Because health insurance is mandatory, each patient pays an amount p for the health insurance plan. The insurance company makes zero profits.[4] Patients receive utility $u = q - \theta T - p$ in each period, where q is the treatment quality they receive, with $q = 1$ if they are treated by senior physicians, while $q = a$ if they are treated by junior physicians.

Timing: The hospital and patients play the following stage game for two periods:

1. The hospital announces whether it is implementing a random or a free-choice system for assigning patients to physicians.
2. Observing the allocation systems, each patient decides which hospital (and in the case of the free-choice system, which physician) to visit.

We assume that all players have a discount factor of 1. We assume that the waiting list will be cleared within the period. First, when physicians are over-demanded, they can stretch their capacity by working overtime to treat all waiting patients within this period. In practice, the German healthcare system guarantees that patients can make an appointment with a specialist within 4 weeks, which can be seen as an example of our assumption. Second, compared to patient's waiting, physician's learning takes a longer time. Therefore, we assume that patient's waiting is a within period event and physician's learning is a cross-period event. Finally, this assumption ensures that no patients are left without treatment at the end of the game. Note that a monopolist will have no incentive to change the allocation system between the two periods. In the following, we study the hospital's payoff, the patients' choices, and the level of social welfare under each system.

2.1 | Random-allocation system

In $t=1$, patients are randomly assigned to a physician. If a patient visits the hospital, she has an expected treatment value of $\frac{1+a}{2}$. Otherwise, the patient receives a treatment value of 0. Given that each patient has to pay the health insurance fee p, every patient will visit a physician. Each patient receives an expected utility $u_1 = \frac{1+a}{2} - p$. The total of all patients' utilities is $\int_0^1 u_1 d\theta = \frac{1+a}{2} - p$.

Each physician treats a quarter of the population in the random-allocation system. Since each physician can serve a quarter of the population, there never is a queue. This allows each physician to earn a payoff of $\frac{p}{4}$, and the total payoff for the hospital amounts to $\pi_1^r = p$.

The social welfare in $t=1$ is

$$w_1^r = \pi_1^r + \int_0^1 u_1 d\theta = \frac{1+a}{2}.$$

In $t=2$, patients are still randomly assigned to a physician. Each physician still treats a quarter of the population. Each physician earns the same payoff as before, and the same is true of the hospital. However, since a junior physician gains seniority due to having treated enough patients in $t=1$, each patient's utility becomes $u_2 = 1-p$. Therefore, the social welfare in $t=2$ is

$$w_2^r = \pi_2^r + \int_0^1 u_2 d\theta = 1.$$

Over the two periods, the hospital earns a total payoff of $\pi^r = 2p$, and the total social welfare is $w^r = w_1^r + w_2^r = \frac{3+a}{2}$.

2.2 | Free-choice system

In the free-choice system, we assume that patients are aware of the seniority of the physicians. There are two reasons for this assumption. First, patients are motivated to identify who is a senior physician since everyone wants to be treated by a superior physician for the same price. Second, a senior physician has an incentive to inform patients of his status since he can earn more if more patients visit him. In practice, patients often rely on word of mouth or consult physician-rating websites before deciding which physician to visit. Similarly, hospitals and insurance providers often disclose information regarding physicians' ability in the hope that competition pressure will increase the quality of service provided by medical practitioners.

In $t=1$, patients prefer to be treated by a senior physician. However, if more than a quarter of the patient population chooses to visit a senior physician, a queue will inevitably form. Therefore, the more a patient dislikes waiting, the less likely she is to wait for a senior physician. Given the index of patients' dislike to wait, results show that a patient with a cutoff index $\hat{\theta}$ is indifferent between waiting for a senior physician or visiting a junior practitioner. In other words, this patient has the same utility, no matter which physician she visits. This implies

$$1 - \hat{\theta}T - p = a - p.$$

A patient with an index smaller than $\hat{\theta}$ always requests treatment from a senior physician. On the other hand, a patient with a higher index always requests care from the junior physician. In a two-period model, we assume that physicians have to clear the waiting list in each period. The expected waiting time T is characterized by two factors. The first factor represents the probability of waiting. Given the total capacity of senior physicians is $\frac{1}{2}$ and $\hat{\theta} > \frac{1}{2}$ patients visit them, each visiting patient has a probability of $\frac{\hat{\theta}-\frac{1}{2}}{\hat{\theta}}$ to be placed on the waiting list, which is the ratio between the excess demand and the demand. The second factor represents the conditional waiting time: the average waiting time for each senior physician, given that there is a queue. Since the length of the waiting queue for two senior physicians is $\hat{\theta}-\frac{1}{2}$, each senior physician faces a queue length $\frac{\hat{\theta}-\frac{1}{2}}{2}$. Physicians treat patients one by one, and each treatment takes one unit of time. If a patient ranks first in the waiting list, the waiting time for her is 0; if a patient ranks last in the waiting list, the waiting time for her is $\frac{\hat{\theta}-\frac{1}{2}}{2}$. Hence, if a patient is waiting for a senior physician, the expected waiting time for her is the half-length of the waiting line, that is, $\frac{\hat{\theta}-\frac{1}{2}}{2}\times\frac{1}{2}$. To calculate the expected waiting time, we multiply the probability of waiting by the conditional waiting time, that is, $T=\frac{\hat{\theta}-\frac{1}{2}}{\hat{\theta}}\times\left(\frac{\hat{\theta}-\frac{1}{2}}{2}\times\frac{1}{2}\right)$. Thus we know that the cutoff index is

$$\hat{\theta}=\frac{1}{2}+2\sqrt{1-a}.$$

Any patient with a dislike-to-wait index higher than the cutoff will choose a random junior physician, whereas any patient with a lower dislike-to-wait index will choose a random senior physician. The cutoff index depends on the quality difference $1-a$: the difference between junior and senior physicians' ability. If the quality difference is sufficiently large $\left(a\le\frac{15}{16}\right)$, all patients will visit the senior physicians although they anticipate an additional burden of waiting. In the case of a smaller quality difference $\left(a>\frac{15}{16}\right)$, some patients will visit the junior physicians. However, more patients will still visit the senior practitioners. In this case, four physicians are active in the market.

2.2.1 | Only the senior physicians are active

In $t=1$, for $a<\frac{15}{16}$, all patients visit a random senior physician. Therefore, each junior physician earns 0 and each senior physician earns $\frac{p}{2}$. The hospital earns a payoff $\pi_1^f=p$. Since the demand for a senior physician exceeds his capacity, there is a queue. Each patient obtains an expected utility $u_1=1-\theta T-p$, where $T=\frac{1}{16}$.

Therefore, the social welfare in $t=1$ amounts to

$$w_1^f=\pi_1^f+\int_0^1 u_1 d\theta=\frac{31}{32}.$$

In $t=2$, the junior physicians do not turn senior since they did not treat enough patients in $t=1$. The quality difference between the two physicians is, therefore, the same as before. Therefore, each physician's payoff, the hospital's payoff, and patients' utilities do not differ from $t=1$.

Over two periods, the hospital earns a total payoff $\pi^f=2\pi_1^f=2p$, and the total social welfare is $w^f=2w_1^f=\frac{31}{16}$.

2.2.2 | All physicians are active

In $t=1$, for $\frac{15}{16} \leq a < 1$, the quality difference between the junior and the senior physicians is small. Patients with a dislike-to-wait index greater than $\hat{\theta}$ visit a junior physician, while those with a lower dislike-to-wait visit a senior physician. Therefore, a junior physician earns $\frac{(1-\hat{\theta})}{2}p$, and a senior practitioner earns $\frac{\hat{\theta}}{2}p$. The hospital earns a payoff $\pi_1^f = p$.

Since the demand for a senior physician is greater than his capacity, patients have to wait. Each patient visiting a senior physician receives an expected utility $u_1 = 1 - \theta T - p$, where $T = \frac{\hat{\theta}-\frac{1}{2}}{\hat{\theta}} \times \left(\frac{\hat{\theta}-\frac{1}{2}}{2} \times \frac{1}{2} \right)$. Since there is no queue for a junior physician, each patient receives an expected utility $u_1 = a - p$.

Therefore, the social welfare in $t=1$ is:

$$w_1^f = \pi_1^f + \int_0^{\hat{\theta}} (1 - \theta T(\hat{\theta}) - p) d\theta + \int_{\hat{\theta}}^1 (a-p) d\theta = a + \frac{(1-a)\left(\frac{1}{2} + 2\sqrt{1-a}\right)}{2}.$$

In $t=2$, the junior physicians retain their status since they did not treat enough patients in $t=1$. Hence, the quality difference in $t=2$ is the same as before. Therefore, each physician's payoff, the hospital's payoff, and patients' utilities are the same as before.

Over two periods, the hospital earns a total payoff $\pi^f = 2\pi_1^f = 2p$, and the total social welfare is

$$w^f = 2w_1^f = 2a + (1-a)\left(\frac{1}{2} + 2\sqrt{1-a}\right).$$

2.3 | Welfare comparison

The monopolistic hospital is indifferent to whether it implements the random-allocation or the free-choice system. Both systems lead to the same payoff. However, when a policy-maker who wishes to maximize social welfare makes this decision, the policy-maker will choose the system based on the difference in ability. Figure 1 illustrates the choice of allocation system by the policy-maker for different junior ability a.

Proposition 1 *When the quality difference between a junior physician and a senior physician is sufficiently great* $\left(1 - a \geq \frac{1}{8}\right)$, *the free-choice system leads to greater social welfare than the random-allocation system. For a small quality difference* $\left(1 - a_1 < \frac{1}{8}\right)$, *the random-allocation system leads to greater social welfare.*

Proof The welfare difference between the two systems is denoted by $\Delta w = w^r - w^f$. Then, we have

$$\Delta w = \begin{cases} \frac{8a-7}{16} & \text{if } a < \frac{15}{16} \\ (1-a)(1-2\sqrt{1-a}) & \text{if } \frac{15}{16} \leq a < 1. \end{cases}$$

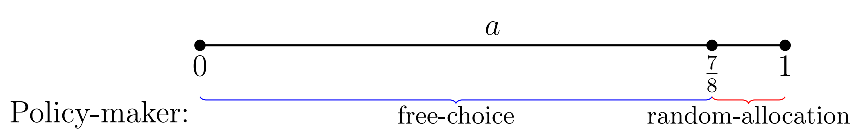

FIGURE 1 Welfare-maximizing allocation under monopoly depending on juniors' ability a. The blue (red) line represents the junior ability for which the free-choice (random) allocation is welfare maximizing [Colour figure can be viewed at wileyonlinelibrary.com]

When $a \leq \frac{7}{8}$, we have $\Delta w \leq 0$ and when $a > \frac{7}{8}$, we have $\Delta w > 0$. □

Since the hospital earns the same payoff under both systems, the difference in social welfare is solely due to the difference in the surplus of patients. Given the quality difference between junior and senior physicians, patients weigh up the waiting cost for seeing a senior physician and the quality difference between the senior and junior physicians to decide which physician to visit. The free-choice system outweighs the random-allocation system because it provides patients with the chance to optimize their choices internally. During the first period under the random-allocation system, some patients are allocated to a junior physician to enable him to achieve senior status by the second period, although they would prefer to see a senior practitioner. In contrast, no patient has to visit a junior physician in the first period under the free-choice system. However, a junior physician cannot improve his ability if he lacks patients in the first period. For a large quality difference, the disutility arising from being randomly assigned to a practitioner under the random-allocation system in the first period outweighs the utility to be gained in the second period from the increase in junior-physicians' learning in the first period.

3 | COMPETITION IN THE MARKET

In a monopolistic market, the monopolist hospital earns the same payoff no matter which system is implemented. Hence, the monopolist would not mind implementing the system that maximizes social welfare. In this section, we show that this is no longer the case when there is competition. To introduce competition, we split the monopolistic hospital with its four physicians into two hospitals: hospital A and hospital B. The hospitals now have two physicians each, one junior and one senior, and compete for patients in the market. To make our analysis comparable to the monopolistic case, we assume that physicians can only treat $\frac{1}{4}$ of the population without patients having to wait. The timing of the stage game is as follows (see Figure 2):

1. In $t = 1$, hospitals simultaneously announce whether they will implement a random-allocation or a free-choice system to assign patients to physicians.
2. In $t = 1$, observing the allocation systems, each patient decides which hospital (and, if free choice, which physician) to visit.
3. In $t = 2$, hospitals simultaneously announce whether they will change their allocation system or not.
4. In $t = 2$, observing the current allocation systems, each patient decides which hospital (and, if free choice, which physician) to visit.

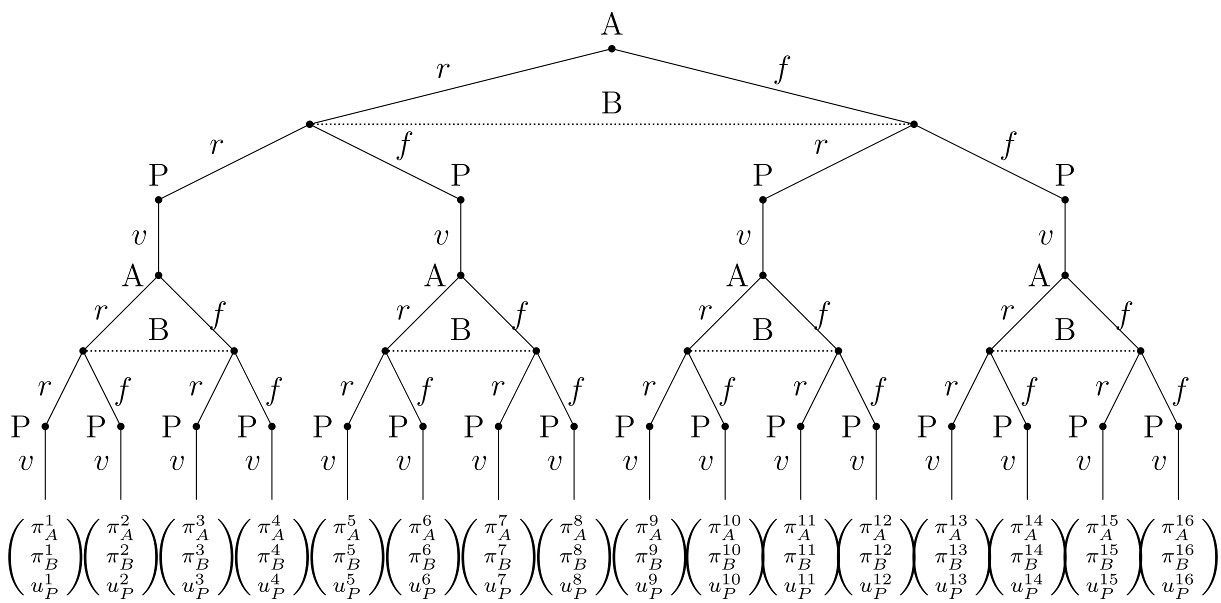

FIGURE 2 Game tree of the competition in the market

The actions and payoffs of players are illustrated in the above game tree, where hospitals A and B choose between a random-allocation system (r) and a free-choice system (f), and patients (P) choose which hospital (and, if free-choice, which physician) visit (v). Hospital B's information set reflects the fact that it chooses its allocation system at the same time as hospital A. The total payoffs over two periods for hospital A, hospital B, and patients are denoted by π_A^s, π_B^s, and u_P^s respectively, where $s \in \{1, 2, \ldots, 16\}$ denotes different combinations of allocation systems.

We pin down the subgame perfect Nash equilibrium of the above game by backward induction. The subgame starts in $t = 2$ when hospitals A and B decide about their allocation systems. We first look at the Nash equilibrium of this subgame, then go back to the game played in $t = 1$ to find the subgame perfect Nash equilibrium of the whole game. Note that in $t = 2$, the junior physician of each hospital could turn senior upon receiving a sufficient number of patients in $t = 1$, and that the number of patients depends on the combination of allocation systems of hospitals in $t = 1$. Hence, to find the Nash equilibrium of the subgame, we have to analyze how the choices of each hospital on the allocation system in $t = 1$ impact the treatment quality of each hospital in $t = 2$.

Given each hospital has two options for its allocation system, four possible combinations of actions arise in $t = 1$. Suppose a hospital chooses system i and the other chooses system j in $t = 1$, where $i, j \in \{r, f\}$, with r denoting the random-allocation system and f the free-choice system. Denote the hospital's payoff in $t = 1$ by π_1^{ij}. For the subgame starting from this system combination, we can find its Nash equilibrium in $t = 2$, where hospitals may choose different allocation systems from their previous ones. Denote the hospital's payoff in the Nash equilibrium of the subgame in $t = 2$ by π_2^{ij}. The total payoff over two periods is denoted by $\pi^{ij} = \pi_1^{ij} + \pi_2^{ij}$. For instance, π^{rf} refers to the total payoffs of the hospital that implements the random-allocation system, while its competitor implements the free-choice system in $t = 1$. By comparing the total payoffs over different combinations of allocation systems in $t = 1$, we determine the subgame perfect Nash equilibrium for the whole game. The four combinations in $t = 1$ are as depicted in Table 1, where the payoffs are the total payoffs on the sequential equilibrium path.

3.1 | Both choose random-allocation systems in $t = 1$

If both hospitals choose the random-allocation system in $t = 1$, patients will randomly visit one hospital. Since each junior physician treats a sufficient number of patients in the first period, they will both achieve seniority for the second period. Since then, there is no quality difference between physicians in $t = 2$, and any combination of allocation systems can be a Nash equilibrium in this subgame. Then, in $t = 2$, each physician serves $\frac{1}{4}$ of the population independent of the system each hospital adopts.

The payoff for each hospital in each period is $\pi_1^{rr} = \pi_2^{rr} = \frac{p}{2}$. Therefore, when both hospitals choose the random-allocation systems in $t = 1$, each will get a total payoff of $\pi^{rr} = p$. The total welfare in the two periods amounts to $w^{rr} = \frac{3+a}{2}$.

3.2 | One random-allocation system and one free-choice system in $t = 1$

Without loss of generality, we assume that hospital A chooses the random-allocation system and hospital B chooses the free-choice system in $t = 1$. Without the need to wait, a patient receives a utility of $\frac{1+a}{2} - p$ if she visits hospital A; a utility of $1 - p$ if she visits the senior physician at hospital B; and a utility of $a - p$ if she visits the junior physician at hospital B. Figure 3 illustrates this situation. Hence, without considering the waiting time, all patients would visit the senior physician at hospital B. As a senior physician can only serve $\frac{1}{4}$ of the patients without delay, there will be a queue.

TABLE 1 Four combinations of hospitals' allocation systems in $t = 1$ and their total payoffs on the sequential equilibrium path

		B	
		Random-allocation	**Free-choice**
A	Random-allocation	π^{rr}, π^{rr}	π^{rf}, π^{fr}
	Free-choice	π^{fr}, π^{rf}	π^{ff}, π^{ff}

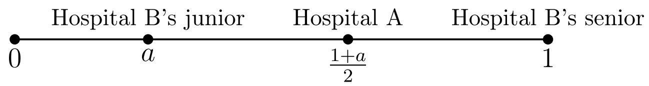

FIGURE 3 Ranking of treatment quality when hospital A chooses a random-allocation system and hospital B chooses a free-choice system

Patients weigh the difference in treatment quality against the waiting cost. When the quality difference between a junior physician and a senior becomes smaller, patients with a high dislike-to-wait may switch to hospital A. When the quality difference becomes even smaller, patients with a high dislike to wait may switch to the junior physician at hospital B.

Lemma 1 *When the quality difference between a senior and a junior physician is large* $\left(a < \frac{15}{16}\right)$, *hospital A receives less than half of the patient population in* $t = 1$; *otherwise, hospital A receives no less than half of the patient population.*

Proof Given the ranking of the treatment quality among hospital A and two physicians at hospital B, hospital A is receiving less than half of the patient population in $t = 1$, implying that no patient waits at hospital A. There are two scenarios: either no patient visits hospital A or some patients visit hospital A, but no one is waiting there.

Scenario 1: no patient visits hospital A. Recall that a patient's expected utility is $u = q - \theta T - p$. Let us focus on the patient with the highest dislike-to-wait index. If this patient prefers visiting the senior physician at hospital B with waiting over visiting hospital A without waiting, no patient will visit hospital A. The condition is $1 - 1 \times \frac{1-\frac{1}{4}}{1} \times \frac{1-\frac{1}{4}}{2} - p \geq \frac{1+a}{2} - p$, which implies $a \leq \frac{7}{16}$.

Scenario 2: some patients visit hospital A, but no one is waiting there. Let us focus on the patient who is indifferent to whether she visits the senior physician at hospital B with waiting or visits a random physician at hospital A without waiting. Denote the dislike-to-wait index of the patient by $\hat{\theta}$. Then, we have $1 - \hat{\theta} \times \frac{\hat{\theta}-\frac{1}{4}}{\hat{\theta}} \times \frac{\hat{\theta}-\frac{1}{4}}{2} - p = \frac{1+a}{2} - p$. We get $\hat{\theta} = \frac{1}{4} + \sqrt{1-a}$. Patients with a dislike-to-wait index smaller than $\hat{\theta}$ will thus visit the senior physician at hospital B. All other patients will visit a random physician at hospital A. This requires $\hat{\theta} < 1$, which gives us $a > \frac{7}{16}$. No patient waiting at hospital A implies less than half of the patients visiting hospital A. Therefore, it must be that $\hat{\theta} > \frac{1}{2}$, which give us $a < \frac{15}{16}$.

Hence, when the quality difference between a senior and a junior physician is large $\left(a < \frac{15}{16}\right)$, hospital A receives less than half of the patient population in $t = 1$.

We prove the second part by contradiction. When the quality difference is small $\left(a \geq \frac{15}{16}\right)$, let us suppose hospital A receives less than half of the patient population. Given that the capacity of hospital A is $\frac{1}{2}$, it implies no patient waits at hospital A. Since the quality of treatment at hospital A is higher than that of the junior physician at hospital B, no patient visits the junior physician at hospital B. Hence, patients choose between visiting hospital A without waiting and visiting the senior physician at hospital B. Denote by $\hat{\theta}$ the dislike-to-wait index of the patient who is indifferent between visiting hospital A or visiting the senior physician at hospital B. From what we discussed above, we know $\hat{\theta} = \frac{1}{4} + \sqrt{1-a}$. Patients with a dislike-to-wait index bigger than $\hat{\theta}$ visit hospital A. Hospital A receiving less than half of the patient population implies $1 - \hat{\theta} > \frac{1}{2}$, which gives us $a < \frac{15}{16}$. This contradicts the condition $a \geq \frac{15}{16}$. Hence, when the quality difference is small $\left(a \geq \frac{15}{16}\right)$, hospital A receives no less than half of the patient population. □

Lemma 2 *If hospital A receives less than half of the patient population in* $t = 1$*, hospital A earns less than hospital B over two periods that is,* $\pi^{fr} > p > \pi^{rf}$*; otherwise, hospital A earns more than hospital B over two periods that is,* $\pi^{fr} < p < \pi^{rf}$.

Proof When hospital A receives less than half of the patient population in $t = 1$, we know hospital A earns less than hospital B in $t = 1$, that is, $\pi_1^{fr} > \frac{p}{2} > \pi_1^{rf}$. In addition, the junior physician at hospital A remains junior in $t = 2$ because the random allocation system does not give the junior physician more than a quarter of the patient population to treat. Since the treatment quality of hospital A is higher than that of the junior physician at hospital B, no one waiting at hospital A implies that the junior physician at hospital B receives no patient in $t = 1$. Given that there is no change in physician's treatment quality in $t = 2$ and hospital A earns less than hospital B in $t = 1$, choosing the free-choice system is the dominant strategy in $t = 2$. Hence, the Nash equilibrium of the subgame in $t = 2$ is that both hospitals choose the free-choice systems, which gives us $\pi_2^{fr} = \frac{p}{2} = \pi_2^{rf}$. Therefore, we have $\pi^{fr} > p > \pi^{rf}$.

When hospital A receives no less than half of the patient population in $t = 1$, we know hospital A earns no less than hospital B in $t = 1$, that is, $\pi_1^{fr} \leq \frac{p}{2} \leq \pi_1^{rf}$. In addition, the junior physician at hospital A will become senior in $t = 2$ because the random allocation system gives its junior physician more than a quarter of the patient population to treat. In contrast, the junior physician at hospital B will not become senior in $t = 2$ because hospital B uses the free-choice system, and it receives less than half of the patient population in $t = 1$. Since hospital A has two senior physicians in $t = 2$ and hospital B only has one, hospital A will be indifferent towards the allocation systems; however, for hospital B, choosing the random-allocation system will be the dominant strategy if $a \geq \frac{71}{72}$ and choosing the free-choice system will be the dominant strategy if $\frac{15}{16} < a < \frac{71}{72}$, which consists of the Nash equilibria of the subgame played in $t = 2$. Since hospital A has two senior physicians and hospital B only has one in $t = 2$, hospital A will receive more than half of the patient population in $t = 2$, no matter which Nash equilibrium is played in $t = 2$, which implies $\pi_2^{fr} < \frac{p}{2} < \pi_2^{rf}$. Hence, we have $\pi^{fr} < p < \pi^{rf}$. □

3.3 | Both free-choice systems in $t = 1$

As shown in Section 2.2, when $a < \frac{15}{16}$, no patient visits a junior physician; when $\frac{15}{16} \leq a < 1$, some patients visit junior physicians, but no patient is willing to wait for a junior physician.

Therefore, the payoff for each hospital in $t = 1$ is $\pi_1^{ff} = \frac{p}{2}$. Since no junior physician becomes senior in $t = 2$, the Nash equilibrium of this subgame is that both hospitals choose free-choice systems if $a < \frac{15}{16}$ and both hospitals choose random-allocation systems if $a \geq \frac{15}{16}$, which is demonstrated by Lemma 1. Hence, each hospital $\pi_2^{ff} = \frac{p}{2}$ in $t = 2$. Overall, each hospital earns $\pi^{ff} = p$.

4 | SUBGAME PERFECT NASH EQUILIBRIUM AND WELFARE

Proposition 2 *When the quality difference between a junior physician and a senior physician is large* $\left(a < \frac{15}{16}\right)$*, the subgame perfect Nash equilibrium of the competition is that both hospitals choose the free-choice system. When the quality difference is small* $\left(a \geq \frac{15}{16}\right)$*, the subgame perfect Nash equilibrium of the competition is that both hospitals choose the random-allocation system.*

Proof From the analysis above, we know that $\pi^{fr} = p = \pi^{rf}$ when both hospitals choose either the random-allocation system or the free-choice system. Furthermore, combining Lemma 1 and Lemma 2, we have $\pi^{fr} > p > \pi^{rf}$ for $a < \frac{15}{16}$ and $\pi^{fr} < p < \pi^{rf}$ for $a \geq \frac{15}{16}$ (see Table 2).

Therefore, when $a < \frac{15}{16}$, the subgame perfect Nash equilibrium is that both hospitals choose the free-choice system, and when $a \geq \frac{15}{16}$, the subgame perfect Nash equilibrium is that both hospitals choose the random-allocation system. □

The policy-maker aims to maximize social welfare when choosing an allocation system.[5] Since the competition scenario is identical to equally splitting the monopolistic hospital in two, the policy-maker in the competitive market makes the same choice as in the monopolistic market.

Proposition 3 *When the quality difference between a junior and a senior physician is large or small* $\left(a \leq \frac{7}{8}\right.$ *or* $\left.a \geq \frac{15}{16}\right)$, *the competitive hospitals choose the allocation system that maximizes social welfare. However, when the quality difference is intermediate* $\left(\frac{7}{8} < a < \frac{15}{16}\right)$, *the random-allocation system maximizes social welfare, while the hospitals prefer the free-choice system.*

Proof By putting Propositions 1 and 2 together, we find that for $a \leq \frac{7}{8}$, the policy-maker and the hospitals all prefer the free-choice system; for $a \geq \frac{15}{16}$, the policy-maker and the hospitals prefer the random-allocation system; and for $\frac{7}{8} < a < \frac{15}{16}$, the policy-maker prefers the random-allocation system, but the hospitals prefer the free-choice system. □

The deviation of market equilibrium from the socially optimal choice stems from market imperfection. The market is imperfect in the sense that a patient's choice may create a positive or negative externality for other patients. Under a free-choice system, the marginal patient whose visit leads to a junior physician becoming senior generates a positive externality for all other patients in the next period. When there are patients waiting for a physician, any additional visiting patient will impose a negative externality on other waiting patients. Given these externalities, when the quality difference between a junior and a senior physician is intermediate, the competition between hospitals creates a Prisoner's Dilemma. Choosing a free-choice system is a dominant strategy although both hospitals choose a random-allocation system is the socially optimal. Figure 4 summarizes our results.

TABLE 2 Subgame perfect Nash equilibria under competition

		B	
		Random-allocation	**Free-choice**
A	Random-allocation	p, p	π^{rf}, π^{fr}
	Free-choice	π^{fr}, π^{rf}	p, p

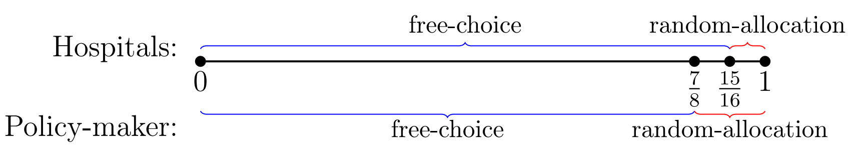

FIGURE 4 Choice of allocation system by hospitals and policy-maker for different junior ability a [Colour figure can be viewed at wileyonlinelibrary.com]

This result implies that in a competitive market in which junior physicians need to improve their skills by treating a sufficient number of patients, maximum social welfare is not attained if hospitals choose their allocation systems when staffed with junior physicians of intermediate ability.

5 | DISCUSSION AND EXTENSIONS

Our model allows us to mirror healthcare markets under different designs. In reality, physicians face an increasing and convex cost function to provide service when their capacities are limited. When we assume that the price of treatment is high enough to cover the highest cost of treatment, physicians are willing to provide the treatment, which implies all patients will be treated. In addition, physicians do not choose treatment quality directly in our model because this is decided by the number of patients that physicians has treated in the past. As a result, the cost of treatment does not affect treatment quality.

We assume for the sake of simplicity that junior physicians either turn senior in the second period or not. However, learning might be a continuous rather than a discrete process. Even treating only a few patients in the first period may lead to a small improvement of ability in the second period. In the Appendix, we show that the externalities still remain under these conditions, although to a smaller extent. Our model can thus be applied under conditions of different healthcare designs, different learning processes, and different types of patient disutility resulting from not being treated immediately: the results are qualitatively robust across all these variations.

We allow senior physicians to put as many patients on the waiting list as demand allows. Regulation, however, may restrict the maximum number of patients that a senior physician has on his waiting list. The interval of ability for which the random-allocation system maximizes social welfare increases once such regulation is implemented. This is because the opportunity costs of learning decrease in the first period under the additional regulation.

Patients' awareness of the physicians' seniority is an important assumption. If patients are not aware of quality difference between junior and senior physicians, a free-choice system is the same as a random-allocation system. A further research question could be what the optimal allocation system would be under different quality gaps between juniors and seniors if patients are not fully aware of physicians' seniority.

Another feature of our model is that although we have simplified it to exclude price discrimination, our insights also hold when senior physicians discriminate. Senior physicians discriminate by charging a higher fee for the higher quality of their care. These extra charges are not covered by mandatory health insurance in most European countries. This charging more than the health insurance fee is often referred to as “balance billing”. As is the case in most European countries, with a public and a private sector with balance billing, we assume in our model that senior physicians cannot discriminate between patients when charging fees. Thus, senior physicians set their extra charges so that the marginal patient they can treat is indifferent to whether she is treated by a senior or a junior physician.

6 | CONCLUSION

Policy-makers aim to improve quality in healthcare markets. A key question is how junior physicians can acquire the experience necessary to become senior physicians. We present a learning model in which hospitals each have a junior and a senior physician. Hospitals can choose between two systems how to allocate their patients to their two physicians: a free-choice system and a random-allocation system. While the free-choice system ensures that patients are allocated according to their preferences, it does not ensure that a junior physician gains sufficient experience to become a senior physician. In contrast, the random-allocation system ensures that junior physicians treat a sufficient number of patients. In this case, however, patients are not matched to physicians according to their preferences. We show that in a monopolistic market, social welfare is higher (lower) under the free-choice system than the random-allocation system when the quality difference between junior and senior physicians is sufficiently large (small). In a competitive market, patients choose between hospitals. If the difference in ability between junior and senior physicians is sufficiently small or sufficiently large, the hospitals will choose the allocation system that maximizes social welfare. When junior physicians are of intermediate ability, the random-allocation system maximizes social welfare, but hospitals will prefer to implement the free-choice system. Therefore, a policy intervention is necessary to maximize social welfare when the quality difference between junior and senior physicians is intermediate.

ACKNOWLEDGMENT

We would like to thank Jan Boone, Heiko Karle, Mathias Kifmann, and the seminar participants at the European Health Economics Workshop for very helpful comments and suggestions.

Open Access Funding provided by Eidgenossische Technische Hochschule Zurich.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ORCID

Christian Waibel https://orcid.org/0000-0002-0162-7644

ENDNOTES

1 The Dutch health minister complained that patients' freedom to choose their own healthcare provider leads to higher expenses in the healthcare system. See https://www.dutchnews.nl/news/2019/03/healthcare-market-has-gone-too-far-says-dutch-health-minister/, accessed on March 27, 2019.

2 This assumption does not change our qualitative results. A lower cutoff for turning senior favors the random-allocation system.

3 In our model, hospitals are physician owned. Hence, a hospital's payoff is what its physicians earn.

4 In reality, the insurance company does not transfer all the insurance fees to hospitals. However, this kind of redistribution does not affect patients' choice of physicians or the total social welfare. Hence, we make such a simple assumption to focus on allocating systems to social welfare.

5 We assume that the policy-maker cannot discriminate between hospitals when implementing an allocation system. Hence the policy-maker can put either a random-allocation system or a free-choice system for both hospitals into practice.

REFERENCES

Beitia, A. (2003). Hospital quality choice and market structure in a regulated duopoly. *Journal of Health Economics*, *22*, 1011–1036.

Brekke, K., Gravelle, H., Siciliani, L., & Straume, O. (2014). Patient choice, mobility and competition among health care providers, In R. Levaggi & M. Montefiori (Eds.), *Health care provision and patient mobility* (pp. 1–26), Springer-Verlag.

Brekke, K., Nuscheler, R., & Straume, O. (2007). Gatekeeping in health care. *Journal of Health Economics*, *26*, 149–170.

Brekke, K., Siciliani, L., & Straume, O. (2008). Competition and waiting times in hospital markets. *Journal of Public Economics*, *92*, 1607–1628.

Brekke, K., Siciliani, L., & Straume, O. (2011). Hospital competition and quality with regulated prices. *The Scandinavian Journal of Economics*, *113*, 444–469.

Buckley, N. J., Cuff, K., Hurley, J., McLeod, L., Nuscheler, R., & Cameron, D. (2012). Willingness-to-pay for parallel private health insurance: Evidence from a laboratory experiment. *Canadian Journal of Economics*, *45*, 137–166.

Chalkley, M., & Malcomson, J. (1998). Contracting for health services with unmonitored quality. *The Economic Journal*, *108*, 1093–1110.

Cooper, Z., Gibbons, S., Jones, S., & McGuire, A. (2011). Does hospital competition save lives? Evidence from the English NHS patient choice reforms. *The Economic Journal*, *121*, 228–260.

Cuff, K., Hurley, J., Mestelman, S., Muller, A., & Nuscheler, R. (2012). Public and private health-care financing with alternate public rationing rules. *Health Economics*, *21*, 83–100.

Cutler, D. M. (1996). *Public policy for health care (Technical Report)*. National Bureau of Economic Research.

Gong, Q. (2017). *Physician learning and treatment choices: Evidence from brain aneurysms (Working Paper)*. University of Pennsylvania.

Hehenkamp, B., & Kaarbèe, O. M. (2020). Location choice and quality competition in mixed hospital markets. *Journal of Economic Behavior & Organization*, *177*, 641–660.

Howard, D. (2006). Quality and consumer choice in healthcare: Evidence from kidney transplantation. *Journal of Economic Analysis & Policy*, *5*, 1349.

Huesmann, K., & Mimra, W. (2021). *The streetlight effect: Focusing on unimportant attributes*. Mimeo.

Ma, C. (1994). Health care payment systems: Cost and quality incentives. *Journal of Economics and Management Strategy*, *3*, 93–112.

Narayanan, S., & Manchanda, P. (2009). Heterogeneous learning and the targeting of marketing communication for new products. *Marketing Science*, *28*, 424–441.

Prendergast, C. (1999). The provision of incentives in firms. *Journal of Economic Literature*, *37*, 7–63.

Rebitzer, J. B., & Taylor, L. J. (2011). Extrinsic rewards and intrinsic motives: Standard and behavioral approaches to agency and labor markets. In *Handbook of labor economics* (Vol. 4, pp. 701–772). Elsevier.

Sá, L., Siciliani, L., & Straume, O. (2019). Dynamic hospital competition under rationing by waiting times. *Journal of Health Economics*, *66*, 260–282.

Sobel, J. (2000). Economists' models of learning. *Journal of Economic Theory*, *94*, 241–261.

Tay, A. (2003). Assessing competition in hospital care markets: The importance of accounting for quality differentiation. *The RAND Journal of Economics*, *34*, 786–814.

Varkevisser, M., Van der Geest, S., & Schut, F. (2012). Do patients choose hospitals with high quality ratings? Empirical evidence from the market for angioplasty in the Netherlands. *Journal of Health Economics*, *31*, 371–378.

Young, J. Q., Ranji, S. R., Wachter, R. M., Lee, C. M., Niehaus, B., & Auerbach, A. D. (2011). 'July effect': Impact of the academic year-end changeover on patient outcomes: A systematic review. *Annals of Internal Medicine*, *155*, 309–315.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Li, X., & Waibel, C. (2021). Patients' free choice of physicians is not always good. *Health Economics*, *30*(11), 2751–2765. https://doi.org/10.1002/hec.4407