Modeling Patients' Illness Perception and Equilibrium Analysis of Their Doctor Shopping Behavior

Fengfeng Huang

School of Management and Economics, University of Electronic Science and Technology of China, China, ffhuang@uestc.edu.cn

Pengfei Guo

College of Business, City University of Hong Kong, Kowloon, Hong Kong, penguo@cityu.edu.hk

Yulan Wang*

Faculty of Business, The Hong Kong Polytechnic University, Kowloon, Hong Kong, yulan.wang@polyu.edu.hk

When a patient's illness perception is inconsistent with a doctor's diagnosis, she may seek opinions from multiple doctors without referrals, a behavior called doctor shopping. In this study, we model and derive patients' optimal doctor shopping decisions. After each visit, patients update their beliefs about their health status following the Bayes' rule. We show that the patients' doctor-shopping decisions are critically affected by the diagnosis accuracy, the relative value of identifying a severely ill patient, and the cost per visit. We examine how the patients' doctor shopping behavior affects social welfare from two aspects, namely, an objective one that accesses whether doctor shopping improves the judgment accuracy regarding the patient's health status, and a subjective one concerning whether doctor shopping relieves patients' anxiety. We find that allowing patients to conduct doctor shopping exacerbates the system congestion (*congestion effect*), but it can help those patients who have decided to join obtain a higher reward (*reward effect*). There exists a diagnosis accuracy threshold above which allowing doctor shopping incurs a welfare loss and below which allowing doctor shopping improves welfare. Moreover, this diagnosis accuracy threshold increases as patients become more pessimistic or hold more diverse initial illness perceptions. The objective welfare maximization prefers a higher doctor shopping rate than the subjective welfare maximization does only when the value of identifying a severely ill patient is high enough, which may help explain why doctor shopping is encouraged for the critical illness such as cancer.

Keywords: Doctor Shopping; Healthcare; Queueing Strategy

*Corresponding Author.

1 Introduction

Patients often seek opinions from multiple doctors during a single illness episode without referrals; such behavior is referred to as *doctor shopping* by Kasteler et al. (1976). Doctor shopping is a commonly-observed phenomenon. Nearly 40% of patients in the government out-patient departments in Hong Kong have doctor shopping experience (Lo et al., 1994). In the United States, the prevalence of doctor-shopping among breast cancer patients is 20% (Morrow et al., 2009). The overall prevalence rate of doctor shopping is 18% in Canada (Macpherson et al., 2001) and 23% in Japan (Sato et al., 1995). According to Payne et al. (2014); Sansone and Sansone (2012), the reasons why patients conduct doctor shopping include confirmation of a diagnosis/treatment, dissatisfaction with an initial consultation, unfulfilled needs, a desire for additional information, hopes for a change of a diagnosis, high risks for major morbidity or mortality, and treatment complications and adverse effects.

Despite the prevalence of doctor shopping, the empirical evidence on its implications is mixed. Some believe that it reduces the risk of misdiagnosis and provides patients with better treatment options; see, e.g., Althabe et al. (2004). Others argue that it may lead to patient confusion, resources waste, and a higher risk of in-hospital complications, especially when there is no informed reconciliation of conflicting opinions (Chang et al., 2013; Shmueli et al., 2017). It has not yet reached an conclusion of whether doctor shopping truly improves the quality of care in the general medical practice. This motivates us to develop a thorough understanding of doctor shopping and its effects on the patients and the system performance. In this work, we examine the patients' doctor-shopping behavior when they seek the diagnosis service. We are particularly interested in the following research questions:

(1). How shall we model and derive patients' doctor shopping behavior?

(2). How do the system parameters, such as the diagnosis accuracy, patients' initial illness perceptions, and the features of the illness (namely, the values/losses of being identified/misidentified), affect the equilibrium outcomes, such as the patients' joining incentives, the doctor shopping rate, and the social welfare?

(3). How shall the doctor shopping behavior be regulated from the viewpoint of welfare maximization?

To model patients' doctor shopping behavior, we note that patients often hold prior beliefs (often biased) about their medical conditions, which are called their *illness perceptions* (Weinman and Petrie, 1997). When the diagnosis result provided by the doctor significantly differs from a patient's belief, cognitive dissonance arises, and the patient may resort to doctor-shopping to alleviate the dissonance (Donkin et al., 2006). We consider a stylized public system that offers the diagnosis service. Patients arrive according to a Poisson process. They exhibit similar symptoms but are heterogeneous in their illness perceptions. Each patient's illness perception measures her subjective *belief* about the likelihood that she is severely ill, which is not an informative indicator of her true health status (Petrie et al., 2007). After performing a diagnosis, the doctor dismisses the patient if the diagnosis result is negative (indicating that the patient has a mild health problem) and refers her for treatments if the result is positive (indicating that the patient has a severe health problem). The diagnosis, however, is imperfect. Based on the diagnosis result, the patient updates her belief following a Bayes rule and decides whether to take the advice and leave the diagnostic system or to conduct doctor shopping.

The patient's doctor shopping decision is an optimal stopping problem: it involves comparing the expected reward of stopping immediately with that can be achieved by paying another visit, wherein the future expected reward is associated with another round of stopping-now or continuing-doctor-visiting decision and so on. There exist two illness perception thresholds in equilibrium, an upper one and a lower one. Patients join the diagnostic system when their *initial* illness perceptions are above the lower threshold, and they leave once their *updated* illness perceptions are above the upper threshold or below the lower threshold. We show that any patient whose illness perception falls in between the two thresholds will leave the system when the diagnosis results are the same in two successive visits. Based on the stopping decisions of the patients who join the system, we classify them into three categories, obedient, stubborn, and diagnosis-dependent. The obedient patients always follow the doctor's advice, whereas the stubborn patients are assertive that they have severe health problems and insist on referrals regardless of the diagnosis result;[1] both types visit once and leave. Doctor shopping only occurs among diagnosis-dependent patients. We then derive the equilibrium number of times that a patient visits during one illness episode and the system effective arrival rate. We show that the expected visiting times of the patients are jointly determined by the diagnosis accuracy, the relative value of identifying a severely ill patient, and the cost per visit.

After characterizing the patients' doctor shopping decisions, we examine how doctor shopping affects the system performance from two aspects, namely, the objective welfare and the subjective welfare. The former is concerned with to what degree doctor shopping helps achieve more accurate judgment on the patients' health status while the latter is concerned with how much doctor shopping helps relieve patients' anxiety. We show that

[1] The existence of stubborn patients has been documented in literature; see, e.g., O'Donnell (2000); Webb and Lloyd (1994); Armstrong et al. (1991); Virji and Britten (1991) These studies find that doctors are generally quite responsive to patients' request for referrals even when the diagnosis result is negative.

when doctor shopping is prohibited, there exists an optimal effective arrival rate that the welfare maximization aims to achieve. This optimal effective arrival rate increases with diagnosis accuracy (i.e., the quality of the diagnosis service). When doctor shopping is allowed, we show that the social planner sets a relatively low price so that both the low- and the high-illness-perception diagnosis-dependent patients have incentives to conduct doctor shopping, thereby maximizing the objective social welfare.

Allowing patients to conduct doctor shopping affects the overall system performance in a twofold way. On the one hand, it results in a higher system congestion and thus increases patients' cost per visit, namely the *congestion effect*, which induces more patients to balk. On the other hand, the opportunity of doctor shopping can make the patients who have decided to join obtain a higher reward, namely the *reward effect*, which incentivizes more patients to join. Their net effect determines the patient's optimal joining-or-balking decision and whether to conduct doctor shopping.

We find that when a patient population is more pessimistic (i.e., when patients' initial illness perceptions are more likely to take large values), allowing patients to conduct doctor shopping attains a higher welfare improvement if the diagnosis accuracy is not high. This is because under this situation, the congestion effect induced by allowing doctor shopping is dominated by the corresponding reward effect. However, if the diagnosis accuracy is very high, the increased patient reward from doctor shopping is marginal compared with the exacerbated congestion. In this situation, the more pessimistic patient population incurs a higher welfare loss. When patients' views towards their health conditions are more diverse among the patient population (i.e., when patients' initial illness perceptions are more *mean-preserving* spreading), allowing patients to conduct doctor shopping attains a higher welfare improvement. The effective arrival rate of new patients is smaller and the equilibrium size of stubborn patients in the population is larger as patients' initial illness perceptions become more diverse. Under this situation, the reward improvement from doctor shopping surpasses the corresponding exacerbated congestion. Consequently, allowing patients to conduct doctor shopping leads to a higher welfare improvement. Our results show that for any given patient initial illness perception distribution, there exists a diagnosis accuracy threshold above which allowing doctor shopping incurs a welfare loss and below which allowing doctor shopping improves welfare. Moreover, this diagnosis accuracy threshold increases as patients become more pessimistic or hold more diverse initial illness perceptions.

We further find that when the value of identifying a severely ill patient increases, the welfare improvement from doctor shopping increases and fewer patients balk. Thus, the effective arrival rate of new patients increases. The objective welfare maximization prefers a higher doctor shopping rate than the subjective welfare maximization does only when the value of identifying a severely ill patient is high enough. This may help explain why doctor shopping is encouraged for the critical illness such as cancer (which has a huge impact on the patients' life) (Garcia et al., 2018).

The remainder of this paper is organized as follows. Section 2 reviews the related literature. Model setup is discussed in Section 3. In Section 4, we analyze the patients' doctor shopping behavior and derive the welfare-maximizing pricing strategy. Section 5 compares the social welfare achieved by allowing doctor shopping with that by prohibiting it. Concluding remarks are provided in Section 6. All of the proofs are relegated to the online Appendix.

2 Literature Review

Our study is closely related with studies on the provision of diagnosis service. The works of Wang et al. (2010) and Alizamir et al. (2013) consider balancing congestion and diagnosis accuracy. Wang et al. (2010) consider a multi-server queueing system and demonstrate that with dual concerns over accuracy and congestion, increasing capacity might increase congestion. Alizamir et al. (2013) model the diagnostic process as a sequence of imperfect tests to determine a customer's type. They find that the provider should continue to perform the diagnosis as long as its current belief that the customer is of a given type falls into an interval. The studies of Pac and Veeraraghavan (2010) and Dai et al. (2017) consider the credence nature of the diagnostic service, where the service is provided by experts and customers lack sufficient knowledge to evaluate the service quality. Pac and Veeraraghavan (2010) examine the issue of expert cheating (over- and under-provision). They show that congestion concerns mitigate expert cheating and that high prices can signal honest diagnoses and can reduce the system congestion. Dai et al. (2017) consider the credence nature of the physicians' decisions on ordering imaging tests and find that insurance coverage is the key driver of physicians' over-provision.

Our research is also closely related with those studies concerning the speed-quality trade-off in service systems. De Vericourt and Zhou (2005) study a call-routing system in which customers call back when their problems are not completely resolved. They show that the optimal routing policy is of threshold type. Chan et al. (2014) consider a state-dependent queuing network and find that speedup may actually exacerbate congestion due to the increased need for rework. Yom-Tov and Mandelbaum (2014) examine some time-varying staffing policies of an Erlang-R queueing model where customers may return to service several times during the sojourn time. Guo et al. (2016) investigate the relationships between the payment schemes (fee-for-service and bundled payment), the doctors' service-time decisions, and the patients' readmission. They show that the bundled payment performs better in maximizing patient welfare and reducing readmission while the fee-for-service scheme is preferred in reducing system congestion.

In the aforementioned two streams of studies, service quality is solely determined by the service provider; customers are passive receivers and and readmission is driven by the service failure. In contrast, here, we take into account patients' active decision making based on the behavioral medicine research (Leventhal et al., 1984; Leventhal and Cameron, 1987), where the primary driving force of a patient revisiting the service system is due to the discrepancy between his/her own belief about the self health status and the doctor's diagnosis result. The above-mentioned studies are concerned with the "objective" implications of customer revisits on the system performance including system workload and congestion, whereas our study is concerned with not only the objective aspect regarding the judgment accuracy but also the subjective aspect regarding relieving patients' anxiety by allowing doctor shopping.

Our work is related to those studies that consider the sequential process of customers soliciting the service-related information and examine customers' active decision making on the service provider selection. Hassin and Roet-Green (2018) study the equilibrium joining strategies of customers in a system with two parallel servers, in which each arriving customer chooses a server to inspect its queue length with an inspection cost and then decides whether to join this queue or to further inspect the other queue. They show that in many cases even when servers are symmetric, customers' equilibrium joining strategy is not the monotonic threshold-type[2] but contains cascades (or "holes"). Yang et al. (2019) investigate customers' search behavior when they are confronted with a large collection of vertically differentiated and congested service providers. Customers conduct a costly sequential search to resolve uncertainty about service providers' quality and queue length and follow certain stopping rules to decide which one to join. They find that reducing either the search cost or the customer arrival rate may increase the average waiting time in the system and decrease customer welfare. Different from the above studies, we consider an unobservable queue and examine patients' continuing-doctor-shopping-or-stopping-now decisions by characterizing the evolution of patients' beliefs upon obtaining diagnosis results. Similar to Yang et al. (2019), we find that patients' individual rational decisions may lead to welfare loss due to the increased system congestion. In this sense, our work is also related to Cui et al. (2019) that considers customers' rational retrial decisions in a steady-state queueing system and finds that the retrial option can hurt consumer welfare.

The study on the provision of health service under the framework of queueing games is also related. Green et al. (2006) consider the patient scheduling and appointment acceptance

[2]The monotonic threshold-type joining strategy is the one in which customers join if the queue length is below a threshold and balk otherwise.

in the diagnostic facilities. Green and Savin (2008) model the medical appointment system as a single-server queueing system in which customers have a state-dependent probability of "no-show" and "no-show" customers may rejoin. They demonstrate that the queueing models are helpful in physician practices for managing the "patient panel" and reducing appointment backlogs. Guo et al. (2014) study a two-tier health care system with both a public and a private system and find that expanding the public service capacity could increase congestion for the overall patients; that is, the Downs-Thomson paradox occurs in the two-tier health care system. Qian et al. (2017) examine how to subsidize patients and move them away from the public waiting queue into the private health care system so as to better utilize the service capacities. Shi et al. (2020) research into trade-offs between managing a heavy clinical workflow and conducting valuable medical research on new diagnostic tests and develop a decomposition algorithm.

3 Model Setup and Preliminaries

Consider a public medical diagnostic system that consists of a large (or infinite) number of doctors who are able to provide the same kind of diagnosis service. Each doctor can be considered as a server and their service can be treated as homogeneous. This is reasonable because in a public health care system, medical resources are allocated fairly. For example, the Hong Kong's Hospital Authority allocates resources under "same service, same funding" principle.[3] The doctors' service times are independent and identically distributed exponential random variables with rate μ. The diagnostic system charges a price $f \geq 0$ per each diagnosis service.

A stream of patients exhibiting similar symptoms arrive at each doctor according to a Poisson process with rate Λ. Patients decide whether to join or to balk upon their arrival. Those who decide to join may visit the system multiple times, that is, engaging in doctor shopping. Patients are served by each doctor on the First-Come First-Served (FCFS) basis. Let α_b denote the prevalence of the disease, that is, the probability of a patient being stricken with a severe health problem. Each patient (she) holds an initial *illness perception* (or belief) towards her illness condition, that is, the probability that she believes herself to be severely ill, denoted as α. Patients' illness perceptions αs are independently drawn from a distribution supported on the interval $(0,\ 1)$, with the probability density function (PDF) and the cumulative distribution function (CDF) denoted as $\phi(\cdot)$ and $\Phi(\cdot)$, respectively. Patients' illness perceptions are subjective (Petrie et al., 2007) and hence do not necessarily reflect their true health status.

[3] See `http://www.legco.gov.hk/yr13-14/english/panels/hs/papers/hs0120cb2-671-5-e.pdf`.

The diagnosis helps determine a patient's health status/type denoted by t: being positive and having a severe health problem ($t = 1$) or being negative and having a mild health problem ($t = 0$). After performing the diagnosis, the doctor dismisses the patient if the diagnosis result is negative (denoted by $s = 0$) or recommends her for treatments if the diagnosis result is positive (denoted by $s = 1$). Diagnostic errors can occur. Given a patient's true type $t = i$, the probability that her diagnosis result is $s = j$ is expressed as $q(s = j|t = i)$, where $i, j \in \{0, 1\}$. For brevity, denote $q(s = j|t = i)$ by q_{ij}. Clearly, $q_{10} + q_{11} = 1$ and $q_{01} + q_{00} = 1$. When $i = j$, the diagnosis result is correct. We assume that $\frac{1}{2} < q_{11} = q_{00} = q < 1$, where q is an indicator of the diagnosis accuracy. The condition $q > \frac{1}{2}$ ensures that the likelihood of correct diagnosis is larger than 50%. When $i \neq j$, the diagnosis result is false. The doctor either misidentifies a patient with mild conditions ($t = 0$) as severe ($s = 1$), i.e., a false positive (*type I error*), or misidentifies a patient with severe conditions ($t = 1$) as mild ($s = 0$), i.e., a false negative (*type II error*). Given a patient's true type $t = i$ ($i \in \{0, 1\}$), she gains a reward V_i if she is correctly identified and suffers a loss L_i if misidentified. We assume that $\min\{q_{11}V_1 - q_{10}L_1, q_{00}V_0 - q_{01}L_0\} > 0$. That is, the diagnostic information (with the possibility of misdiagnosis taken into account) is valuable to patients regardless of their true health status.

Let $X_n = x_n$ be the posterior belief of a patient about her health status who has joined the system and paid n ($n \in \boldsymbol{N}$) visits, where $0 < x_n < 1$. Conditional on $X_n = x_n$, the patient decides whether to conduct doctor shopping one more time with a belief of obtaining a positive diagnosis result ($s = 1$) with probability $P(s = 1|x_n)$ and a negative one ($s = 0$) with probability $P(s = 0|x_n)$, which can be derived as

$$P(s = 1|x_n) = q_{01}(1 - x_n) + q_{11}x_n, \tag{1}$$

$$P(s = 0|x_n) = q_{00}(1 - x_n) + q_{10}x_n. \tag{2}$$

If the patient conducts doctor shopping one more time, she further updates her belief according to the Bayes' rule based on the $(n+1)$th diagnosis result. The posterior belief X_{n+1} thus takes one of the following two values: $x_{n+1} = g_1(x_n)$ if the diagnosis result is positive ($s = 1$) or $x_{n+1} = g_0(x_n)$ if the diagnosis result is negative ($s = 0$), where

$$g_1(x_n) := \frac{q_{11}x_n}{q_{11}x_n + q_{01}(1 - x_n)}; \tag{3}$$

$$g_0(x_n) := \frac{q_{10}x_n}{q_{10}x_n + q_{00}(1 - x_n)}. \tag{4}$$

It can be easily shown that both $g_1(x_n)$ and $g_0(x_n)$ are increasing with x_n.

After obtaining a diagnosis, each patient decides whether to leave the system or to rejoin and seek for more opinions (i.e., conduct doctor shopping). As such, the *effective* arrival rate to the system consists of those from both the newly joined patients and doctor-shopping patients who have visited some doctor(s) and decided to revisit again. Since doctors are homogeneous, we can consider a representative doctor to examine the system performance. We assume that the arrival of doctor-shopping patients to the representative doctor follows a Poisson process, and hence the aggregate arrival process (from newly-joined and doctor-shopping patients) is also Poisson. Let λ denote the patient's effective arrival rate to the representative doctor. The patient's mean sojourn time (i.e., waiting time plus service time) per visit at the representative doctor, denoted by w, is thus $w = 1/(\mu - \lambda)$. Hereafter, by "waiting time" we mean the sojourn time. The patient pays the price f per each visit and incurs a waiting cost that is proportional to her waiting time in the system with a unit-time cost c. Accordingly, the expected cost that a patient incurs in each visit is

$$C_p := f + cw = f + c/(\mu - \lambda). \tag{5}$$

Let N denote the number of times that a patient visits the system in one illness episode. Undoubtedly, N is a random variable. The effective arrival rate then can be expressed as

$$\lambda = \Lambda E[N].$$

Note that the expected number of times that a patient visits the system $E[N]$ is a critical indicator of the system congestion, and unnecessary doctor-shopping leads to a waste of medical resources.

As our focus is patients' doctor shopping behavior, we shall consider the following two aspects regarding social welfare: to what extent it improves the chance of correctly identifying patient types and the associated reward, and to what extent it benefits the patients in terms of relieving anxiety and the perceived reward. Accordingly, we name the former the *objective* (labeled o) social welfare and the latter the *subjective* (labeled s) social welfare. The associated objective reward at the nth visit shall be based on the true probability that the patient is ill, which is updated based on an unbiased prior, i.e., the prevalence of the disease α_b, denoted as Y_n; while the corresponding associated subjective reward shall be based on an individual patient's posterior illness perception, which is updated based on her initial subjective illness perception $X_0 = \alpha$, i.e., X_n. Taking the first visit as an example, the updated illness perception X_1 (i.e., the believed probability of being ill) takes the value $g_1(\alpha)$ if the diagnosis result is positive and $g_0(\alpha)$ otherwise. In contrast, Y_1, the true probability of the patient being ill shall be either $g_1(\alpha_b)$ or $g_0(\alpha_b)$.

4 Analysis

In this section, we first construct the patient's diagnosis-seeking decision as an optimal stopping problem. Next, we investigate the patients' optimal stopping decision. We then derive the public diagnostic system's optimal pricing decision of maximizing social welfare. Here, we consider social welfare from two aspects, an objective one and a subjective one. We also consider a benchmark scenario in which doctor shopping is prohibited to facilitate the understanding of the doctor shopping effect.

4.1 Patient's Optimal Stopping Problem

In this section, we assume that the price per each visit f is given and derive the patients' optimal decisions, including whether or not to seek the diagnostic service, when to stop doctor shopping if joining the diagnostic system, and moreover, the expected number of times a patient visits in an illness episode.

A patient makes her joining-or-balking and continuing-doctor-visiting-or-stopping-now decisions by comparing the reward of stopping at the current state with the expected reward that can be achieved by continuing the visiting process and incurring the related cost. This can be formulated as an optimal stopping problem. Let $v(x_n)$ denote the patient's expected reward arising from her optimal stopping decision given her current belief $X_n = x_n$ ($n \in \boldsymbol{N}$). Then, her expected reward from continuing the visit can be expressed as $E\left[v(X_{n+1})|x_n\right]$. If the patient stops at the current state, she ends with an illness perception $X_n = x_n$ and an expected reward $r(x_n)$. Consequently, the optimality equation can be written as

$$v(x_n) = \max\left\{r(x_n), E\left[v(X_{n+1})|x_n\right] - C_p\right\}, \tag{6}$$

where C_p is the expected cost associated with a visit as stated in (5).

Note that the patient' decision of stopping now is associated with two potential outcomes– leaving the system with being reassured as mildly ill or leaving the system by considering herself as severely ill and seeking treatments. Recall that a type $t = i$ patient gains a reward V_i if she is correctly identified and suffers a loss L_i if misidentified, $i = 0, 1$. Then, her expected rewards of being identified as mildly and severely ill are $(1 - x_n)V_0 - x_nL_1$ and $x_nV_1 - (1 - x_n)L_0$, respectively. The patient then chooses the leaving outcome that yields the higher reward, i.e.,

$$r(x_n) = \max\left\{x_nV_1 - (1 - x_n)L_0,\ (1 - x_n)V_0 - x_nL_1\right\}.$$

Consequently,

$$r(x_n) = \begin{cases} (1 - x_n)V_0 - x_nL_1, & \text{if } 0 < x_n < \widehat{\alpha}, \\ x_nV_1 - (1 - x_n)L_0, & \text{otherwise}, \end{cases} \tag{7}$$

where

$$\widehat{\alpha} = \frac{V_0 + L_0}{V_0 + V_1 + L_0 + L_1}. \tag{8}$$

A lower threshold $\widehat{\alpha}$ corresponds to a larger $(V_1 + L_1)/(V_0 + L_0)$, indicating a higher *relative value of identifying a severely ill patient.* Each leaving outcome, according to the self-regulatory compliance theory in behavioral medicine research, represents the patient's "lay diagnosis" of her health condition; see Leventhal et al. (1984) for a systematic review of the compliance theory. Then, (7) indicates that by the patient's "lay diagnosis", she is mildly ill when she holds a belief $x_n < \widehat{\alpha}$ but severely ill if $\widehat{\alpha} \leq x_n < 1$. We can further show the following result:

Lemma 1. *The expected reward arising from the patient's optimal stopping decision* $v(x_n)$ *stated in* (6) *is convex.*

Let S denote the set of patients' illness perceptions X_n's, $n = 0, 1, 2...$ at the *nth* visit, with which patients will not join/revisit the system once her illness perception falls into this set. Note that when $n = 0$, a patients holds the initial illness perception $X_0 = \alpha$ and will not join the system if $\alpha \in S$. In the behavioral medicine research (Leventhal et al., 1984; Leventhal and Cameron, 1987), it is found that a patient' non-compliance occurs when there exist discrepancies between her own view of the self health condition and that from the doctor. We thus assume that if a patient joins the system, she follows the doctor's advice and leaves the system whenever her own "lay diagnosis" is the same as the doctor's diagnosis; that is, if a patient holds an illness perception $X_n = x_n \in (0, \widehat{\alpha})$ (resp. $X_n = x_n \in [\widehat{\alpha}), 1]$, she then leaves the system whenever the diagnosis result is negative (resp. positive); that is, the updated posterior belief stated in (3) and (4) shall satisfy

$$\begin{cases} g_0(x_n) \in S, & \text{if } 0 < x_n < \widehat{\alpha}, \\ g_1(x_n) \in S, & \text{otherwise}, \end{cases} \quad \text{and } \widehat{\alpha} \notin S. \tag{9}$$

We can show that (9) holds as long as the expected cost associated with each visit $C_p = f + cw$ stated in (5) falls into a moderate range and satisfies [4]

$$\max\{g_0(\widehat{\alpha})(1-\widehat{\alpha}),\ [1 - g_1(\widehat{\alpha})]\widehat{\alpha}\} \leq \frac{C_p}{V_0 + V_1 + L_0 + L_1} < (q_{11} - q_{10})\widehat{\alpha}(1-\widehat{\alpha}).$$

In our optimal stopping problem, both the expected reward arising from the optimal stopping decision $v(x_n)$ and the expected reward of stopping immediately $r(x_n)$ highly hinge on the patient's illness perception $X_n = x_n$, which depends closely on her initial subjective

[4]See the online Appendix for the detailed derivation of the conditions that ensure (9) holds.

illness perception $X_0 = \alpha$; that is, the patient's optimal stopping decision is based on the *subjective reward*. Her *objective reward*, however, shall be based on the true probability of being ill Y_n, which is updated based on the unbiased prior α_b. Let

$$\mathcal{C} := \left[\sqrt{\left(\frac{q_{00}q_{10}}{q_{00}-q_{10}}\right)^2 + q_{00}q_{10}\widehat{\alpha}(1-\widehat{\alpha})} - \frac{q_{00}q_{10}}{q_{00}-q_{10}}\right](V_0 + L_0 + V_1 + L_1).$$

Based on Lemma 1, we obtain the following proposition regarding the patient's optimal stopping decision.

Proposition 1. *The optimal stopping set S takes the following form:*

$$S = \{X_n | 0 < X_n \leq \underline{\alpha} \text{ or } \overline{\alpha} \leq X_{n+1} < 1, \; n = 0, 1, 2, ...\},$$

where $\underline{\alpha} < \widehat{\alpha} < \overline{\alpha}$ and

$$\underline{\alpha} = \begin{cases} \frac{q_{01}^2(L_0+V_0)+(1+q_{01})C_p}{q_{01}^2(L_0+V_0)+q_{11}^2(L_1+V_1)-(q_{11}-q_{01})C_p} & \text{if } 0 < C_p < \mathcal{C}, \\ \frac{q_{01}(L_0+V_0)+C_p}{q_{11}(L_1+V_1)+q_{01}(L_0+V_0)} & \text{otherwise}; \end{cases} \tag{10}$$

$$\overline{\alpha} = \begin{cases} \frac{q_{00}^2(L_0+V_0)-(1+q_{00})C_p}{q_{00}^2(L_0+V_0)+q_{10}^2(L_1+V_1)-(q_{00}-q_{10})C_p} & \text{if } 0 < C_p < \mathcal{C}, \\ \frac{q_{00}(L_0+V_0)-C_p}{q_{00}(L_0+V_0)+q_{10}(L_1+V_1)} & \text{otherwise}. \end{cases} \tag{11}$$

A patient whose illness perception falls into the interval $(\underline{\alpha}, \overline{\alpha})$ leaves the system whenever two successive diagnosis results are the same.

Proposition 1 shows that a patient stops visiting the system whenever her illness perception is either below a lower threshold $\underline{\alpha}$ or above an upper threshold $\overline{\alpha}$. This indicates that patients makes their joining-or-balking and continuing-doctor-visiting-or-stopping-now decisions by following a double-threshold stopping policy. Proposition 1 also indicates that if a patient's initial illness perception satisfies $0 < \alpha \leq \underline{\alpha}$, she will not join the system as she perceives no need to take medical action. However, when the patient's initial illness perception is above the upper threshold, namely when $\overline{\alpha} \leq \alpha < 1$, she still needs to join the system because diagnosis is required prior to treatment. Once a patient joins the system, she keeps on visiting doctors until her posterior illness perception falls into the stopping set S.

The two thresholds highly hinge on the relative magnitudes of the reward that a patient is correctly identified (V_1 and V_0) and the potential loss that she is misidentified (L_0 and L_1), the diagnosis accuracy q_{00} and q_{11}, and the cost associated with each visit C_p. Note that the magnitudes of V_1, V_0, L_0 and L_1 depend on the illness type and the related clinical development, and the diagnosis accuracy is usually determined by the available technologies and qualifications of doctors. They are unlikely to be changed in the short run. However, the system can easily adjust its price per each visit f to alter the patient's visiting cost C_p and thus to manipulate her visiting incentive. We can further show that when the expected cost per visit C_p is relatively high, i.e. when $C_p > \mathcal{C}$, the optimal continuing-doctor-visiting or stopping-now decision rule characterized by the two thresholds $\underline{\alpha}$ and $\overline{\alpha}$ exactly follows the *One-Step Look-Ahead* (OSLA) rule;[5] that is, at each decision point, the patient only compares the current reward obtained by stopping immediately with that by conducting one more time of doctor shopping.

To facilitate the analysis, denote

$$\underline{\alpha}_b := \frac{\overline{\alpha} q_{01}}{\overline{\alpha} q_{01} + (1-\overline{\alpha}) q_{11}}, \ \overline{\alpha}_b := \frac{\underline{\alpha} q_{00}}{\underline{\alpha} q_{00} + (1-\underline{\alpha}) q_{10}},$$
$$\alpha_s := \frac{\overline{\alpha} q_{00}}{\overline{\alpha} q_{00} + (1-\overline{\alpha}) q_{10}}, \ \text{and } \eta := \frac{\underline{\alpha}(1-\overline{\alpha})}{\overline{\alpha}(1-\underline{\alpha})}.$$

We first provide the following lemma that summarizes the relationships among the above thresholds.

Lemma 2. *η increases with C_p, the expected total cost per visit. Moreover, when $C_p < \mathcal{C}$, $\left(\frac{q_{10}}{q_{11}}\right)^2 \leq \eta < \frac{q_{10}}{q_{11}}$, under which $\underline{\alpha} < \underline{\alpha}_b \leq \overline{\alpha}_b \leq \overline{\alpha} < \alpha_s$; when $C_p \geq \mathcal{C}$, $\frac{q_{10}}{q_{11}} \leq \eta < 1$, under which $\underline{\alpha}_b \leq \underline{\alpha} < \overline{\alpha} \leq \overline{\alpha}_b < \alpha_s$.*

We then examine the patient's optimal visiting decision and obtain the following results:

Proposition 2. *For a patient with the initial illness perception $X_0 = \alpha \in (0, 1)$,*

1. *when $\left(\frac{q_{10}}{q_{11}}\right)^2 \leq \eta < \frac{q_{10}}{q_{11}}$,*

   i. *if $\alpha \in (0, \underline{\alpha})$, she never joins the system;*

   ii. *if $\alpha \in (\underline{\alpha},\ \underline{\alpha}_b)$, she visits and then leaves the system if the diagnosis result is negative; otherwise, she conducts doctor shopping;*

   iii. *if $\alpha \in [\underline{\alpha}_b,\ \overline{\alpha}_b]$, she visits the system once and then follows whatever the advice the doctor gives;*

   iv. *if $\alpha \in (\overline{\alpha}_b,\ \alpha_s)$, she visits and then leaves the system if the diagnosis result is positive; otherwise, she conducts doctor shopping;*

   v. *otherwise, she always requests for a referral of treatment regardless of the diagnosis result.*

[5]The analysis of the stopping set under the OSLA rule can be found in the proof of Proposition 1 in the online Appendix.

2. *when* $\frac{q_{10}}{q_{11}} \leq \eta < 1$*, the patient behaves exactly the same as described above except that when* $\alpha \in (\underline{\alpha},\ \overline{\alpha}_b]$*, she visits the system once and then follows whatever the advice the doctor gives;*

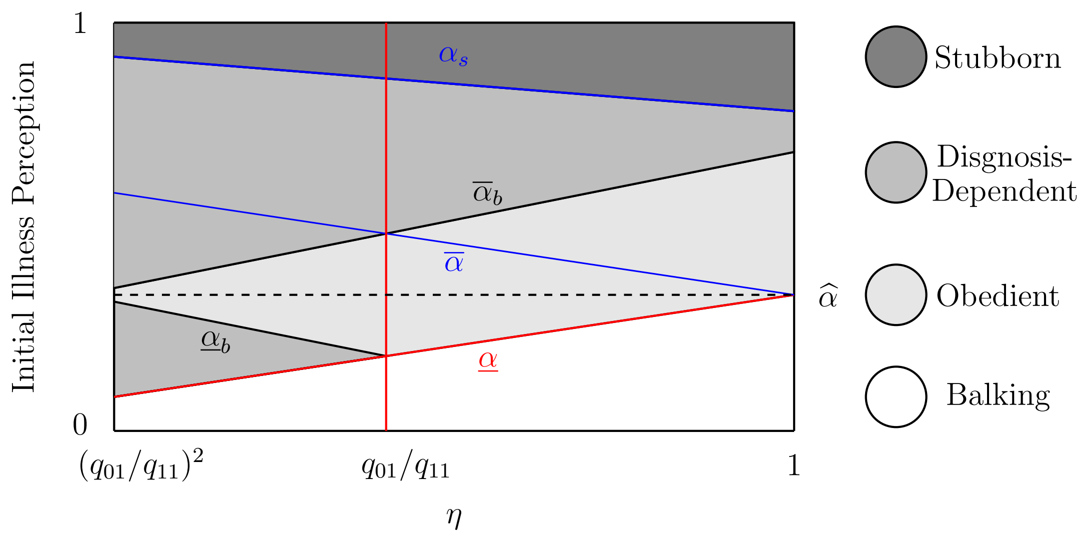


Figure 1: Illustration of the Defined Thresholds and Classified Patient Types

Based on Proposition 2, we can classify the population of patients into the following four types: *balking patients* who never join the system, *diagnosis-dependent patients* who conduct doctor shopping when the diagnosis result deviates from their beliefs, *obedient patients* who always follow the doctor's advice, and *stubborn patients* who require referrals for treatments regardless of the diagnosis results; see Figure 1 for the illustration of patient types. Doctor-shopping behavior occurs only among the diagnosis-dependent patients. From Proposition 2 and illustrated by Figure 1, we can see that when the threshold η is small (i.e., $\eta < \frac{q_{01}}{q_{11}}$), doctor shopping occurs among patients with both relatively low and relatively high illness perceptions ($\alpha \in (\underline{\alpha},\ \underline{\alpha}_b) \cup (\overline{\alpha}_b,\ \alpha_s)$). However, when $\eta \geq \frac{q_{01}}{q_{11}}$, the doctor shopping behavior occurs only among patients with relatively high illness perceptions ($\alpha \in (\overline{\alpha}_b,\ \alpha_s)$) and the OSLA rule is adopted.

When $\left(\frac{q_{01}}{q_{11}}\right)^2 \leq \eta < \frac{q_{01}}{q_{11}}$, let

$$\widetilde{\alpha} := \frac{\underline{\alpha} q_{00}^2}{\underline{\alpha} q_{00}^2 + (1-\underline{\alpha}) q_{10}^2}.$$

Recall that N is the total number of visits taken for a patient to reach the stopping set (and thus leave the diagnostic system) in one illness episode; that is,

$$N = \min\{n : X_n \in S\}.$$

Obviously, $N = 0$ if $X_0 = \alpha \le \underline{\alpha}$. By Propositions 1 and 2, we can then derive the expected number of times that a patient visits the system.

Proposition 3. *Given the expected cost per visit C_p, the expected number of times that a patient visits the system $E[N]$ is as follows:*

1. *when* $\left(\frac{q_{01}}{q_{11}}\right)^2 \le \eta < \frac{q_{01}}{q_{11}}$,

$$\begin{aligned} E[N] =& 1 - \Phi(\underline{\alpha}) + \left(\alpha_b \frac{q_{11}(1+q_{10})}{1-q_{11}q_{10}} + (1-\alpha_b)\frac{q_{01}(1+q_{00})}{1-q_{00}q_{01}}\right)[\Phi(\underline{\alpha}_b) - \Phi(\underline{\alpha})] \\ &+ \left(\alpha_b \frac{q_{10}(1+q_{11})}{1-q_{11}q_{10}} + (1-\alpha_b)\frac{q_{00}(1+q_{01})}{1-q_{00}q_{01}}\right)[\Phi(\overline{\alpha}) - \Phi(\overline{\alpha}_b)] \\ &+ [q_{10}\alpha_b + q_{00}(1-\alpha_b)][\Phi(\widetilde{\alpha}) - \Phi(\overline{\alpha}) \\ &+ \left(\alpha_b \frac{q_{10}(1+q_{10})}{1-q_{11}q_{10}} + (1-\alpha_b)\frac{q_{00}(1+q_{00})}{1-q_{00}q_{01}}\right)[\Phi(\alpha_s) - \Phi(\widetilde{\alpha})], \end{aligned}$$

   where $\overline{\alpha} \le \widetilde{\alpha} < \alpha_s$*;*

2. *when* $\frac{q_{01}}{q_{11}} \le \eta < 1$,

$$E[N] = 1 - \Phi(\underline{\alpha}) + [q_{10}\alpha_b + q_{00}(1-\alpha_b)][\Phi(\alpha_s) - \Phi(\overline{\alpha}_b)].$$

Proposition 3 shows that the patients' expected number of visits in one illness episode $E[N]$ is impacted by the diagnosis accuracy q_{00} and q_{11}, and the illness perception thresholds $\underline{\alpha}$ and $\overline{\alpha}$. We further note that the illness perception thresholds $\underline{\alpha}$ and $\overline{\alpha}$ stated in (10) and (11) of Proposition 1 highly hinge on the magnitude of the relative value of identifying a severely ill patient $(V_1 + L_1)/(V_0 + L_0)$, which decreases in $\widehat{\alpha}$ stated in (8), and the expected cost per visit C_p. Proposition 3 then implies that how many times the patient visits the system is jointly determined by the diagnosis accuracy, the relative value of identifying a severely ill patient and the price per visit f. It is worth mentioning that $E[N]$ can be less than 1 because of patient balking.

So far, we have obtained the expected number of times that a patient visits the system for a given expected waiting time w. However, the expected waiting time w is in turn determined by the expected number of times that a patient visits the system. When the system reaches a stable state, the two numbers shall coincide. Furthermore, we can show the following result:

Proposition 4. *For any given price f, there exists a unique equilibrium effective arrival rate λ.*

We now provide an iterative algorithm to calculate the equilibrium expected waiting time w or, alternatively, the equilibrium effective arrival rate λ as there exists a one-to-one correspondence between them ($w = 1/(\mu - \lambda)$).

Step 0: Set $\lambda^l = 0$ and $\lambda^h = \mu$.

Step 1: Set $\lambda = \frac{\lambda^l + \lambda^h}{2}$. Obtain the thresholds $\underline{\alpha}$ and $\overline{\alpha}$ based on Proposition 1 and $E[N]$ based on Proposition 3.

Step 2: If $\lambda - E[N]\Lambda < 0$, set $\lambda^l = \lambda$; otherwise, $\lambda^h = \lambda$.

Step 3: If $|\lambda - E[N]\Lambda| < 10^{-4}$, stop; otherwise, go to Step 1.

Through the above iteration process, we can then obtain the patients' equilibrium effective arrival rate and the corresponding expected waiting time for any given price f.

4.2 System's Welfare Maximization

In this section, we examine the public diagnostic system's welfare maximization problem by taking into account of patients' doctor shopping behavior. The system cares about the social performance in two aspects, an objective social welfare denoted by $\mathcal{W}_o$ and a subjective social welfare denoted by $\mathcal{W}_s$. The former is concerned with to what extent doctor shopping helps improve the chance of correctly identifying patient types and the related objective reward, while the latter is concerned with to what extent it helps improve patients' perceived feelings in terms of either relieving anxiety and the subjective reward. Below, we first derive these two types of patient rewards.

When patients are allowed to conduct doctor shopping, patients make their optimal stopping decisions according to those stated in §4.1. Recall that a patient reaches the stopping set S after N visits. At the end of the Nth visit, both the true probability Y_N that the patient is ill, which is updated based on the unbiased prior α_b, and her subjective belief X_N that the patient believes herself to be ill, which is updated based on her initial subjective perception α, are uniquely determined by the trajectory of how the patient's illness perception reaches the stopping set. Then, based on (7), we can write the expected *subjective* reward of a patient holding the initial illness perception $X_0 = \alpha$ as $E_{X_N}[r(X_N)\big|X_0 = \alpha]$ and her *objective* reward as $E_{Y_N}[r(Y_N)\big|X_0 = \alpha]$. We can further derive the average subjective reward and the average objective reward per patient, R_s and R_o, as follows:

$$R_s = E_{X_0}\left[E_{X_N}[r(X_N)\big|X_0 = \alpha]\right] \text{ and } R_o = E_{X_0}\left[E_{Y_N}[r(Y_N)\big|X_0 = \alpha]\right].$$

The public diagnostic system decides the price f to maximize the social welfare $\mathcal{W}_i$, $i = o, s$, as follows:

$$\begin{aligned} &\max_f \mathcal{W}_i = \Lambda\left(R_i - E[N]\frac{c}{\mu - \lambda}\right), \\ &\text{s.t.} \quad f \geq 0. \end{aligned} \tag{12}$$

The system's welfare maximization is equivalent to maximizing the patients' average net rewards (average rewards minus the expected total waiting cost), because the payments between the doctor and patients are internalized. Denote the corresponding optimal price by f_i^*. Note that the price f directly impacts the patient's cost per visit C_p. One immediate observation we can withdraw with respect to the two thresholds derived in Proposition 1 is that a higher f leads to a larger $\underline{\alpha}$ but a smaller $\overline{\alpha}$. An increased $\underline{\alpha}$ indicates that a larger proportion of patients choose to balk. Meanwhile, a decreased $\overline{\alpha}$ leads to a smaller α_s, leading to a larger proportion of stubborn patients whose main purpose of visiting is to obtain the referrals. It results in both a lower overall objective patient reward and a lower overall subjective patient reward.

To facilitate the understanding of the doctor shopping effect, we consider a benchmark scenario in which doctor shopping is prohibited, denoted by a superscript "u". Since patients visit the system at most once, the expected number of times a patient visits the system can be easily shown to be $E[N] = 1 - \Phi(\underline{\alpha})$. The average subjective and objective rewards per patient are, respectively,

$$R_s^u = E_{X_0}\left[E_{X_1}[r(X_1)\big|X_0 = \alpha]\right] \text{ and } R_o^u = E_{X_0}\left[E_{Y_1}[r(Y_1)\big|X_0 = \alpha]\right].$$

The corresponding welfare maximization problem can be written as

$$\begin{aligned} \max_f \ & \mathcal{W}_i^u = \Lambda\left(R_i^u - (1-\Phi(\underline{\alpha}))\frac{c}{\mu - (1-\Phi(\underline{\alpha}))\Lambda}\right), \\ \text{s.t.} \ & f \geq 0, \end{aligned} \tag{13}$$

Denote the corresponding optimal price by f_i^{u*}. Define

$$\kappa(x) := (q_{11}V_1 - q_{10}L_1)x + (q_{00}V_0 - q_{01}L_0)(1-x), \text{ where } x \in [0,1], \tag{14}$$

which represents the expected reward of a patient who holds the illness perception x and does not conduct doctor shopping (either due to that doctor shopping is prohibited or because she is obedient). We then have the following result.

Proposition 5. *When doctor shopping is prohibited,*

1. *under the objective social welfare maximization, if there exists a price $\tilde{f} > 0$ such that the effective arrival rate*
$$\lambda_o^u = \mu - \sqrt{\frac{c\mu}{\kappa(\alpha_b)}},$$
then $f_o^{u*} = \tilde{f}$; otherwise, $f_o^{u*} = 0$.

2. *under the subjective social welfare maximization,*

i. when $q_{11}V_1 - q_{10}L_1 \geq q_{00}V_0 - q_{01}L_0$*, if there exists a price* $\hat{f} > 0$ *such that the effective arrival rate*

$$\lambda_s^u = \mu - \sqrt{\frac{c\mu}{\kappa(\underline{\alpha})}}.$$

then $f_s^{u*} = \hat{f}$*; otherwise,* $f_s^{u*} = 0$*;*

ii. otherwise, the optimal price satisfies $0 \leq f_s^{u*} \leq f_o^{u*}$*, and the effective arrival rate* $\lambda_s^u \geq \lambda_o^u = \mu - \sqrt{\frac{c\mu}{\kappa(\alpha_b)}}$.

Proposition 5 shows that when doctor shopping is not allowed, there exists an optimal effective arrival rate λ_j^u ($j \in \{o,\ s\}$) that the welfare-maximizing provider aims to serve via regulating the price. When such rate is not reachable, the system provides the diagnosis service for free to enhance patients' visiting incentives. Note that $\kappa(\cdot)$ stated in (14) increases with the diagnosis accuracy parameters q_{00} and q_{11}. A close look at Proposition 5 then indicates that a higher diagnosis accuracy leads to a larger optimal arrival rate.

We now investigate the scenario in which doctor shopping is allowed and obtain the following result.

Proposition 6. *When doctor shopping is allowed,* $0 < \widehat{\alpha} \leq \alpha_b$ *and*

$$\lambda|_{f=\overline{f}} \leq \overline{\lambda} := \mu - \sqrt{\frac{c\mu}{\kappa(\alpha_b)}\left(1 + \frac{q_{00}q_{10}[q_{10}\alpha_b + q_{00}(1-\alpha_b)]}{(\underline{\alpha}q_{00} + (1-\underline{\alpha})q_{10})^2}\frac{\phi(\overline{\alpha}_b)}{\phi(\underline{\alpha})}\right)}, \tag{15}$$

the optimal price that maximizes the objective social welfare does not exceed $\overline{f}$*, i.e.,* $0 \leq f_o^* \leq \max\{\overline{f},\ 0\}$*, where* $\overline{f} \in R$ *is the solution to* $\eta - q_{01}/q_{11} = 0$*.*

The threshold price $\overline{f}$ stated in Proposition 6 is the one resulting in $\eta = q_{01}/q_{11}$, where η increases with f (see the proof of Proposition 6 in the online Appendix). This implies that $\eta < q_{01}/q_{11}$ whenever $f < \overline{f}$. By Proposition 2, we have that whenever $f < \overline{f}$, doctor shopping occurs among patients with both relatively low and relatively high illness perceptions, whereas when $f \geq \overline{f}$, doctor shopping occurs only among patients with relatively high illness perceptions. Also, recall that a lower threshold $\widehat{\alpha}$ (given in (8)) indicates a higher relative value of identifying an ill patient. Proposition 6 together with Proposition 2 then indicates that when the relative value of identifying an ill patient is high, to maximize the objective social welfare, the optimal price shall be set such that both high- and low-illness-perception diagnosis-dependent patients are encouraged to conduct doctor shopping. Similar to that when doctor shopping is prohibited, Proposition 6 shows that when doctor shopping is allowed, there also exists an optimal effective arrival rate $\bar{\lambda}$ that the social planner aims to achieve. A close look at the expression of $\bar{\lambda}$ in (15) shows that it highly depends on the magnitude of the diagnosis accuracy q_{00} and q_{11}, the prevalence of the disease among patients α_b, and the relative value of identifying a severely ill patient $(V_1 + L_1)/(V_0 + L_0)$, which impact the patient's joining threshold $\underline{\alpha}$. Moreover, the relationship between them is non-monotonic.

However, when doctor shopping is allowed, the subjective social welfare function and the objective social welfare function when conditions stated in Proposition 6 are violated are not well-behaved. We have to resort to the extensive numerical search to find the optimal price and the corresponding system performance.

5 The Impact of Doctor Shopping

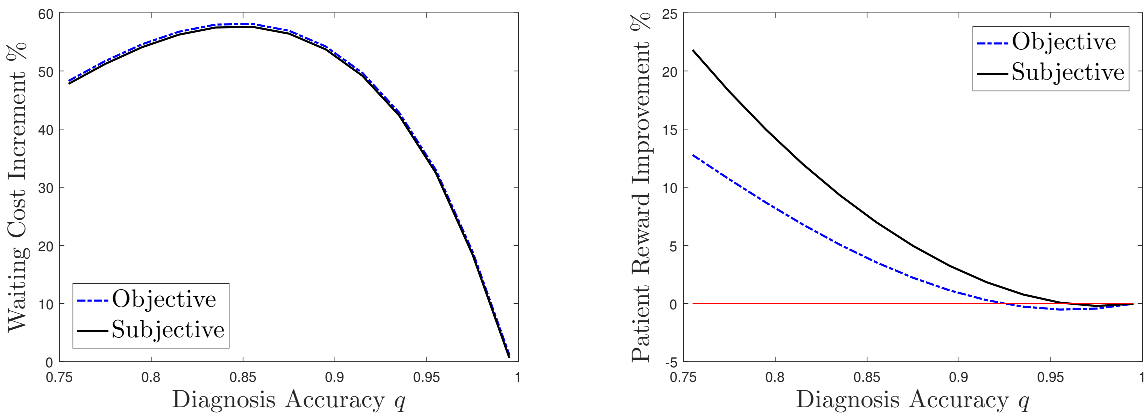
So far, we have analyzed the patients' visiting decision by taking into account of their doctor shopping behavior and the social planer's corresponding welfare maximization problem. We show that the system's service quality reflected by its diagnosis accuracy and the characteristics of the disease reflected by the relative value of identifying a severely ill patient (which impacts the patient's joining incentive) greatly affect the patients' doctor shopping behavior and the system's optimal pricing decision. In this section, we further explicitly explore how they influence the system performance via conducting extensive numerical experiments. The patients' initial illness perceptions (which fall into the interval $[0, 1]$) are drawn from a beta distribution with positive shape parameters a and b, i.e., $Beta(a, b)$, and the diagnosis accuracy $q := q_{00} = q_{11}$.

Consistent with the healthcare literature (Kasteler et al., 1976; Lo et al., 1994; Sato et al., 1995; Macpherson et al., 2001), we define the *doctor-shopping rate* as the proportion of patients who visit more than one doctor in one illness episode. Then, the doctor-shopping rate can be expressed as

$$D_s = \frac{\lambda - (1 - \Phi(\underline{\alpha}))\,\Lambda}{\lambda} \times 100\%.$$

where λ is the effective arrival rate, and $(1 - \Phi(\underline{\alpha}))\,\Lambda$ is the arrival rate of patients who visit the system for the first time. The difference between them is the arrival rate of the doctor-shopping patients. Through a Managed Care Outcomes Project conducted at six health maintenance organizations, Horn et al. (1996) find that among the patients who had one of the five study diseases and needed diagnostic testing, 37.9% were in severe or catastrophic conditions. Moreover, the field study of Petrie and Weinman (2012) finds that prior to testing, patients generally prepare themselves for unfavorable diagnoses with higher illness perception. We thus set the prevalence of the disease among patients $\alpha_b = 0.38$ and consider those beta distributions that satisfy $E[X_0] = \frac{a}{a+b} > \alpha_b = 0.38$ in the numerical experiments.

The other parameter values are set to be $V_1 = V_0 = 120$, $L_1 = L_0 = 80$, $\mu = 4$, $\Lambda = 3$ and $c = 12$.



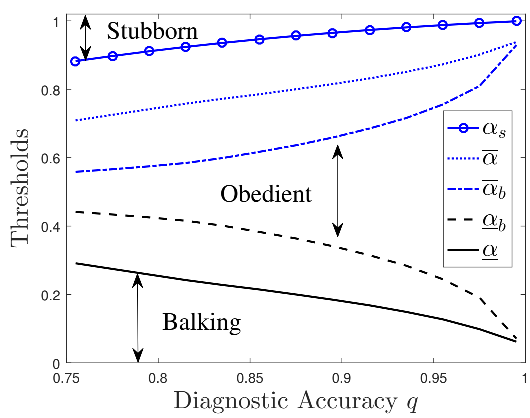
(a) Congestion Effect

(b) Reward Effect

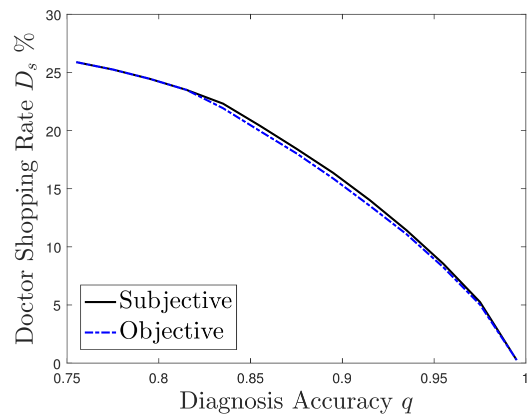
Figure 2: The Impact of Diagnosis Accuracy on the Congestion Effect and Reward Effect of Doctor Shopping: $V_1 = V_0 = 120$, $L_1 = L_0 = 80$, $\mu = 4$, $\Lambda = 3$, $c = 12$, $\alpha_b = 0.38$, $f = 0$, $X_0 \sim Beta(2, 3)$

We first examine how the doctor shopping behavior affects the system congestion and patient rewards compared to those without doctor shopping under a given price in Figure 2, which considers a beta distribution with parameters $(2, 3)$ ($Beta(2, 3)$), the price as $f = 0$, and the diagnosis accuracy q varies from 0.75 to 1. Figure 2(a) depicts the *congestion effect*, which shows that compared to that without doctor shopping, allowing doctor shopping exacerbates system congestion and thus increases patients' waiting cost per visit. Figure 2(b) depicts the *reward effect*, which shows that the opportunity of doctor shopping indicates a higher reward of the patients unless the diagnosis accuracy is very high. Moreover, the reward effect decreases with the diagnosis accuracy. Figure 2(b) further implies that the reward effect in terms of the subjective one concerning relieving patients' anxiety is always higher than that in terms of the objective one concerning improving diagnosis judgment accuracy. Their net effect determines the patient's optimal joining-or-balking decision and whether to conduct doctor shopping.

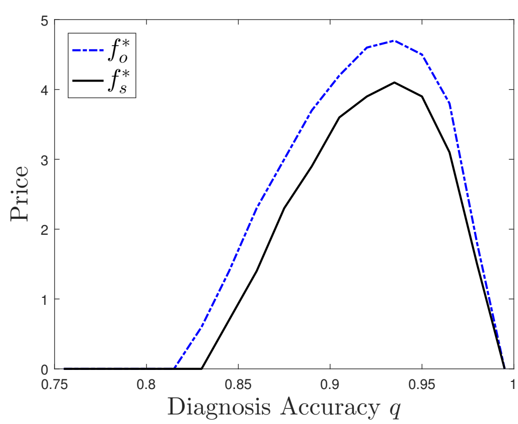
Figure 3 shows the optimal price and the corresponding patient joining behavior in equilibrium when doctor shopping is allowed. Figure 3(a) depicts the critical thresholds stated in Proposition 2 that are used to classify patients' visiting behavior under the optimal objective welfare maximizing price f_o^*. Those under the optimal subjective welfare maximizing price f_s^* exhibit the similar pattern and thus we omit them here. We find that α_s (blue circle line), $\overline{\alpha}_b$ (blue dash-dotted line) and $\underline{\alpha}_b$ (blue dotted line) are increasing with the diagnosis accuracy q, whereas $\underline{\alpha}_b$ (black dash line) and $\underline{\alpha}$ (black solid line) are decreasing with q. Note that a higher diagnosis accuracy q indicates a higher diagnosis service quality and thus the diagnostic result becomes more reliable. Figure 3(a) then shows that as the diagnosis service quality increases, fewer patients balk (due to a lower $\underline{\alpha}$) and thus the system serves more newly-joined patients. Meanwhile, with the increased diagnosis reliability, patients has more confidence in the diagnosis result and thus fewer patients are stubborn (due to a higher α_s). Consequently, more patients are obedient, trust the diagnosis result, and follow the doctor's advice, as reflected by the enlarged interval $[\underline{\alpha}_b,\ \overline{\alpha}_b]$. Patients with the initial illness perception falling into the intervals $(\overline{\alpha}_b, \alpha_s) \cup (\underline{\alpha}, \underline{\alpha}_b)$ are diagnosis-dependent and may doctor-shop. As shown in Figure 3(a), under the optimal price, the length of these two intervals is first enlarged and then narrowed as q increases. This implies that as diagnosis accuracy q increases, the marginal improvement on the patient reward is first increasing and then decreasing.



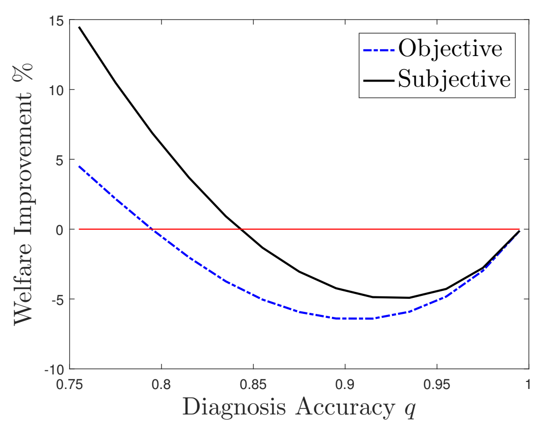
(a) Critical Thresholds



(b) Doctor Shopping Rate



(c) Optimal Price



(d) Welfare Improvement %

Figure 3: The Impact of Diagnosis Accuracy on the System Performance: $V_1 = V_0 = 120$, $L_1 = L_0 = 80$, $\mu = 4$, $\Lambda = 3$, $c = 12$, $\alpha_b = 0.38$, $X_0 \sim Beta(2, 3)$

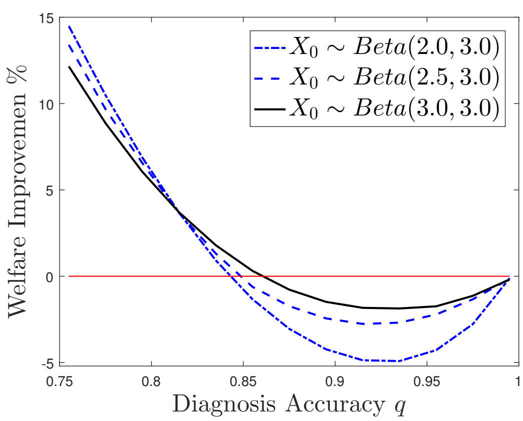
Figure 3(b) shows the resulting equilibrium doctor shopping rates under both the objective and subjective welfare maximization. We can see that both doctor shopping rates decrease with the diagnosis accuracy q, and the one under the objective welfare maximization is slightly lower than the one under the subjective welfare maximization. That is, if the system cares about the patients' subjective rewards associated with their anxiety relieving rather than the objective rewards associated with diagnosis judgment accuracy, it shall encourage more patients to engage in doctor shopping under the former than under the latter. The optimal price that the system shall charge is depicted in Figure 3(c). Figure 3(c) shows that when the diagnosis accuracy q is relatively low, the system shall provide the diagnosis service for free under both the subjective and objective welfare maximization. As q increases, the optimal welfare-maximizing prices f_o^* and f_s^* first (weakly) increase and then decrease, and they both reach the maxima at around $q = 0.94$, with the objective-welfare-maximizing price f_o^* slightly higher than the subjective-welfare-maximizing price f_s^*. This is because a lower price can incentivize more patients to engage in doctor shopping to improve their subjective rewards, as shown in Figure 3(b). Thus, the system is more congested under the subjective welfare maximization than that under the objective welfare maximization.

Figure 3(d) depicts the percentage of welfare improvement from doctor shopping defined as

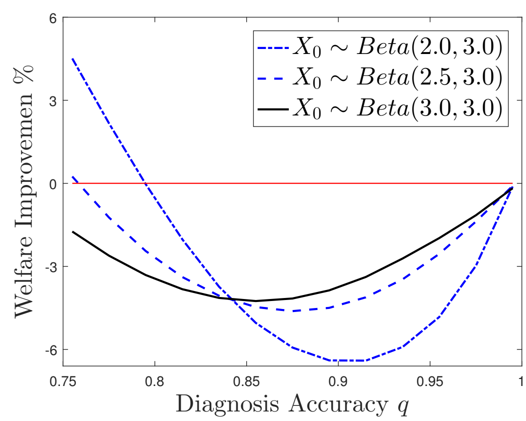
$$\frac{\mathcal{W}_i^* - \mathcal{W}_i^{u*}}{\mathcal{W}_i^{u*}} \times 100\%, i = o, s.$$

It shows that compared to that without doctor shopping, allowing patients to conduct doctor shopping attains a welfare improvement only when the diagnosis accuracy is not high. Moreover, the subjective welfare improvement is higher than the objective one. These results are consistent with those in the medical-related literature concerning whether or not doctor shopping is valuable to diagnosis tests. As noted in the review work by Payne et al. (2014), patients usually believe that second opinions from doctor shopping are valuable, whereas there is no any conclusive evidence of whether doctor shopping improves the quality of patient care. The value of doctor shopping is context-dependent. If doctor shopping stems from *mere* anxiety (which corresponds to the subjective welfare considered in our study), it may lead to patient confusion, resource waste, and a higher risk of in-hospital complications, especially when there is no informed reconciliation of conflicting opinions (Chang et al., 2013; Shmueli et al., 2017). However, doctor shopping is usually believed to improve the quality of care (which corresponds to the objective welfare considered in our study) in radiology and pathology (Payne et al., 2014), because there exist substantial discrepancies in the interpretation of imaging and histopathological diagnosis (i.e., a low diagnosis accuracy). For example, in a study on the value of second opinion for breast cancer patients, Garcia et al. (2018) reported that 43% patients' diagnosis result was changed and concluded that second opinions are valuable for many patients. Our numerical results confirm the above observations in the practice and shed lights on why allowing doctor shopping is beneficial to patients in radiology and pathology but not necessarily so in general care. Specifically, the reality of low diagnosis accuracy in radiology and pathology (changes in diagnosis in 43% of the patients) makes doctor shopping valuable in terms of both the objective and subjective welfare.

Next, we examine how the distribution of the patients' prior belief $X_0 \sim Beta(a, b)$ affect the system performance. Specifically, we vary the distribution in terms of the first-order and the second-order stochastic orders to derive insights. The first-order stochastic ordering measures the "magnitude" of the distribution, where a larger "magnitude" indicates a larger pessimistic patient population. The second-order stochastic ordering measures the distribution's mean-preserving spread-out level, where a higher spread-out level indicates more diverse views of the patients on their health conditions.



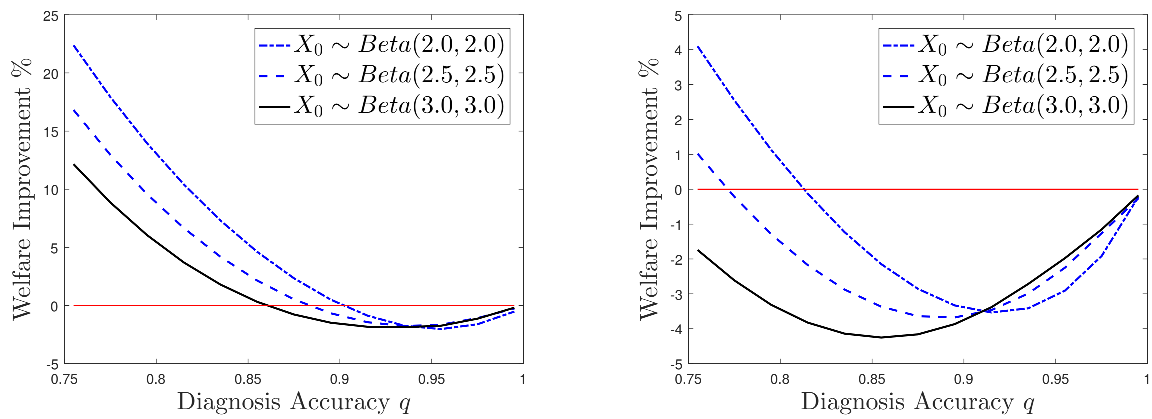
(a) Subjective Welfare



(b) Objective Welfare

Figure 4: The Impact of Allowing Doctor Shopping on the Social Welfare Improvement by Varying Diagnosis Accuracy and the Initial Illness Perception's Beta Distribution in the First-order Stochastic Order: $V_1 = V_0 = 120$, $L_1 = L_0 = 80$, $\mu = 4$, $\Lambda = 3$, $c = 12$, $\alpha_b = 0.38$

We first keep the shape parameter b unchanged but increase the shape parameter a. As such, patients' initial illness perceptions are increasingly likely to take large values, indicating that the patient population becomes more pessimistic. Note that the distribution $Beta(a_1, b)$ first-order stochastically dominates the distribution $Beta(a_2, b)$ if $a_1 > a_2$; see Shaked and Shanthikumar (2007). Figure 4 depicts how the patient population's degree of pessimism together with diagnosis accuracy affect the equilibrium social welfare. It shows that when a patient population becomes more pessimistic (i.e., a larger a), allowing patients to conduct doctor shopping attains a higher welfare improvement if the diagnosis accuracy is not extremely high. This is because when the patient population is more pessimistic, there is a larger proportion of high-illness-perception diagnosis-dependent patients engaged in doctor shopping. When the diagnosis accuracy is not too high, the congestion effect induced by allowing doctor shopping is dominated by the corresponding reward effect. However, if the diagnosis accuracy is extremely high, the increased patient reward from doctor shopping is marginal compared with the exacerbated congestion. In this situation, the more pessimistic patient population incurs a higher welfare loss. Figure 4 demonstrates that there exists a diagnosis accuracy threshold, below which allowing doctor shopping improves social welfare whereas above which allowing doctor shopping hurts social welfare. Moreover, this threshold increases as the patient population becomes more pessimistic (measured by the first-order stochastic order).

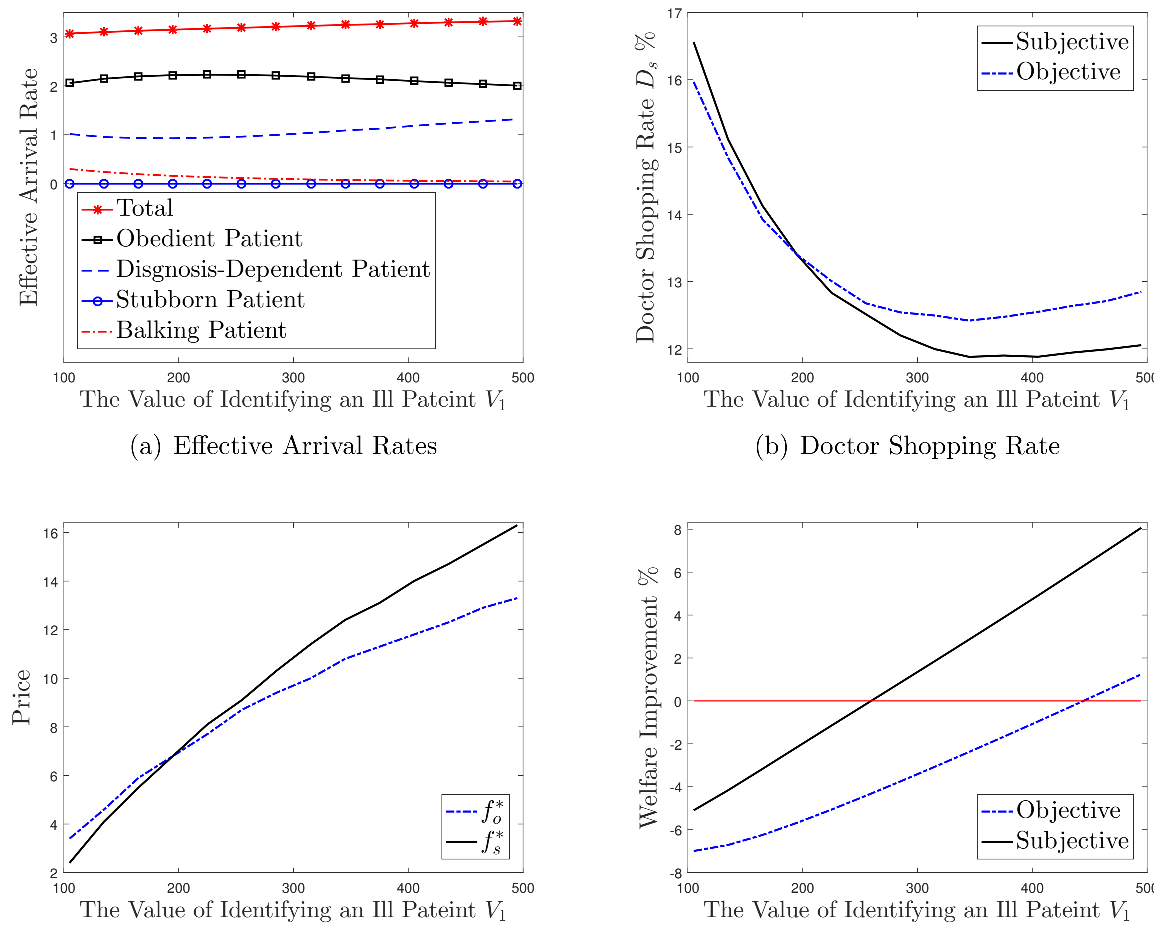


(a) Subjective Welfare Improvement

(b) Objective Welfare Improvement

Figure 5: The Impact of Allowing Doctor Shopping on the Social Welfare Improvement by Varying Diagnosis Accuracy and the Initial Illness Perception's Beta Distribution in the Second-order Stochastic Order: $V_1 = V_0 = 120$, $L_1 = L_0 = 80$, $\mu = 4$, $\Lambda = 3$, $c = 12$, $\alpha_b = 0.38$

We then increase both shape parameters a and b but let $a = b$. As such, the mean of the patients' initial illness perceptions remains the same but the variance becomes smaller as the shape parameters increase, indicating that patients' views towards their health conditions are less diverse. Note that the distribution $Beta(a_1, a_1)$ second-order stochastically dominates the distribution $Beta(a_2, a_2)$ if $a_1 < a_2$; see Shaked and Shanthikumar (2007). Figure 5 shows that as the patient population holds more diverse views (i.e., a smaller a), allowing patients to conduct doctor shopping attains a higher welfare improvement. This is because the effective arrival rate of newly joined patients is smaller and the equilibrium size of stubborn patients in the population is larger as patients' initial illness perceptions become more diverse. Under this situation, the reward improvement from doctor shopping surpasses the corresponding exacerbated congestion. Consequently, allowing patients to conduct doctor shopping leads to a higher welfare improvement. Again, there exists a diagnosis accuracy threshold above which allowing doctor shopping incurs a welfare loss and below which allowing doctor shopping improves welfare. Moreover, this diagnosis accuracy threshold increases as patients hold more diverse initial illness perceptions (measured by the second-order stochastic order).



(a) Effective Arrival Rates

(b) Doctor Shopping Rate

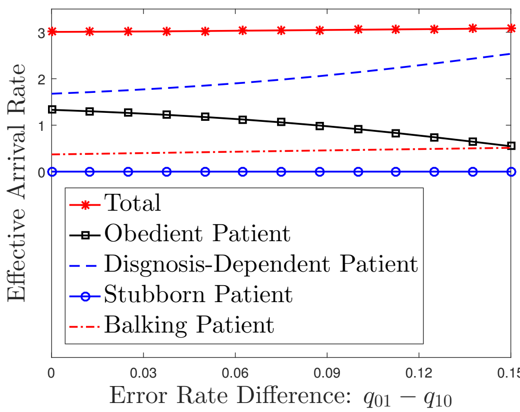
(c) Optimal Price

(d) Welfare Improvement

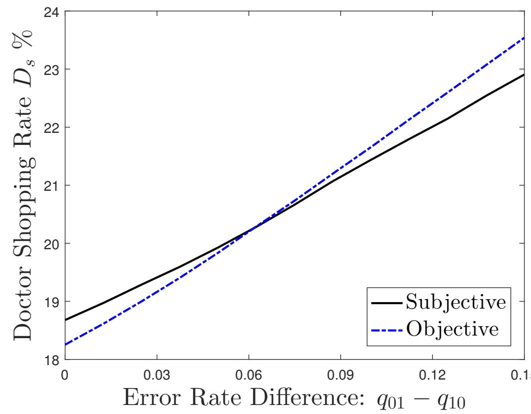
Figure 6: The Impact of the Value of Identifying A Severely Ill Patient on the System Performance: $V_0 = 120$, $L_0 = L_1 = 80$, $\mu = 4$, $\Lambda = 3$, $c = 12$, $\alpha_b = 0.38$, $q = 0.90$, $X_0 \sim Beta(2, 3)$

We further examine how the relative value of identifying a severely ill patient affects the system performance by varying V_1 from 100 to 500 and fixing other parameter values to be $V_0 = 120$, $L_0 = L_1 = 80$, $\mu = 4$, $\Lambda = 3$, $c = 12$, $\alpha_b = 0.38$, $q = 0.90$ and $X_0 \sim Beta(2, 3)$. Note that a larger V_1 indicates a higher value of identifying a severely ill patient. The results are illustrated in Figure 6. Figure 6(a) depicts the equilibrium effective arrival rates under the optimal objective welfare maximizing price f_o^*. Those under the optimal subjective welfare maximizing price f_s^* exhibit the similar pattern and thus are omitted here. In this situation, because the diagnosis accuracy is high ($q = 0.90$), the effective arrival rate of stubborn patients is extremely small (less than 2.75×10^{-4}). As V_1 increases, the relative value of identifying a severely ill patient becomes larger and thus fewer patients balk. Figure 6(a) shows that the effective arrival rate of diagnosis-dependent patients first (weakly) decreases and then increases, whereas that of obedient patients first increases and then (weakly) decreases. Consequently, the total effective arrival rate (from newly-joined patients and doctor-shopping patients) increases with V_1. Accordingly, the doctor shopping rate first decreases and then increases with V_1; see Figure 6(b). Figure 6(b) together with Figure 6(c) show that when the value of identifying a severely ill patient V_1 exceeds a certain threshold, the subjective welfare maximization charges a higher price, which results in a lower doctor shopping rate, than the objective welfare maximization does. Figure 6(d) further shows that the welfare improvement from allowing doctor shopping transforms from a loss to a gain and keeps increasing as V_1 increases. That is, when the value of identifying a severely ill patient is not high (for those diseases that are noncritical), allowing doctor shopping incurs welfare loss as the induced congestion effect dominates the corresponding reward effect. However, when the value of identifying a severely ill patient is high enough, allowing doctor shopping attains welfare improvement. Figure 6(d) together with Figure 6(b) help explain why doctors encourage patients to seek second opinions for those critical diseases that have a huge impact on patients' life (Garcia et al., 2018; Payne et al., 2014). When it comes to the serious health problem like cancer, the value of identifying a severely ill patient is extremely high. Our results show that a higher doctor shopping rate is preferred when maximizing the objective welfare concerning increasing the judgment accuracy compared to that when maximizing the subjective welfare concerning relieving patients' anxiety. This is consistent with that of Manion et al. (2008) which suggest that second opinions shall be mandatory in diagnostic surgical pathology service.

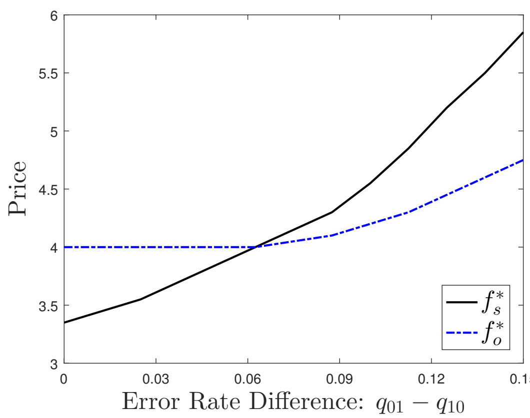
In the above numerical studies, the diagnostic system incurs the two types of diagnostic errors, false positive (*type I error*) and false negative (*type II error*) with the same likelihood; that is, the diagnosis accuracy $q_{00} = q_{11} = q$ and the diagnostic errors $q_{01} = q_{10} = 1 - q$. We now relax this by considering that the likelihoods of false positive and false negative are different and examine how such error rate difference affects the system performance. Specifically, we fix the overall diagnosis accuracy $q := q_{11}\alpha_b + q_{00}(1-\alpha_b) = (1-q_{10})\alpha_b + (1-q_{01})(1-\alpha_b)$ but vary the false positive and false negative error rates q_{01} and q_{10}. We consider the scenario $\frac{1}{2} < q_{00} < q_{11} < 1$, under which the difference between the two diagnostic error rates can be expressed as $q_{01} - q_{10} = q_{11} - q_{00}$, where a larger error rate difference indicates that the diagnostic system is more likely to incur one type of the error over the other. In this scenario, we can first discretize the state of patients' belief (namely, their illness perceptions) into multiple small intervals. We then calculate the expected visiting times for patients in each interval, based on which we can calculate the expected visiting times for the whole patient group. After that, we design an algorithm to iterate between finding the effective arrival rate given the two thresholds and solving the dynamic program to find the two thresholds given the updated effective arrival rate. We refer interested readers to the online Appendix B for the detail.



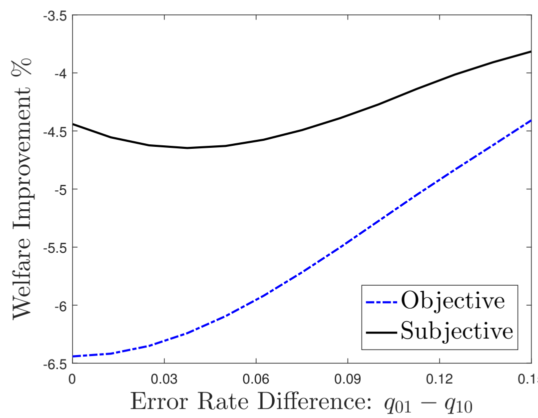
(a) Effective Arrival Rate



(b) Doctor Shopping Rate



(c) Optimal Price



(d) Welfare Improvement %

Figure 7: The Impact of Diagnostic Error Rate Difference on the System Performance: $V_1 = V_0 = 120$, $L_1 = L_0 = 80$, $\mu = 4$, $\Lambda = 3$, $c = 12$, $\alpha_b = 0.38$, $q = 0.90$, $X_0 \sim Beta(2, 3)$

Figure 7(a) depicts the equilibrium effective arrival rates under the optimal objective welfare maximizing price f_o^*. Those under the optimal subjective welfare maximizing price f_s^* exhibit the similar pattern and thus are omitted here. As the overall diagnosis accuracy is high ($q = 0.90$), the effective arrival rate of stubborn patients remains extremely small (less than 1.21×10^{-4}). However, as the error rate difference $q_{01} - q_{10}$ increases, more patients balk and less patients are obedient. Both the effective arrival rate of diagnosis-dependent patients and the total effective arrival rate (from both newly-joined and doctor-shopping patients) (weakly) increase. Consequently, the doctor shopping rate increases with $q_{01} - q_{10}$ as illustrated in Figure 7(b). Figure 7(b) together with Figure 7(c) show that when the error rate difference $q_{01} - q_{10}$ exceeds a certain threshold, the subjective welfare maximization charges a higher price and has a lower doctor shopping rate than the objective welfare maximization does. Figure 7(d) shows that due to the high diagnosis accuracy($q = 0.90$), allowing doctor shopping actually leads to welfare loss under both the objective and subjective welfare maximization. However, a larger error rate difference $q_{01} - q_{10}$ always mitigates the negative effect of allowing doctor shopping on the system welfare under the objective welfare maximization. In contrast, under the subjective welfare maximization, as $q_{01} - q_{10}$ increases, it first exacerbates the negative effect of allowing doctor shopping and enlarges the welfare loss and then dampens such effect and reduces the welfare loss when the error rate difference becomes large enough. The results under the scenario $\frac{1}{2} < q_{11} < q_{00} < 1$ are qualitatively the same.

6 Conclusion

In this study, we model and analyze the doctor shopping behaviors of the patients in a public diagnostic system. Patients are heterogeneous in their illness perceptions (i.e., subjective beliefs of the probability of being in the severe condition). Upon receiving the diagnosis results, they actively decide whether to follow a doctor's advice or to seek second opinions. According to the visiting patterns of the patients, we show that they can be classified into the following four types, balking, obedient, stubborn, and diagnosis-dependent. Doctor shopping occurs only among the diagnosis-dependent patients. We then derive the expected number of times that a patient visits the system in one illness episode and the effective arrival rate to the system.

Our research sheds lights on the implication of doctor shopping on the system performance. On the one hand, allowing patients to conduct doctor shopping exacerbates the system congestion and hence increases patients' waiting cost, which may induce more patients to balk. On the other hand, the opportunity of doctor shopping can improve patient rewards and hence enhance patients' joining incentive. Their net effect determines the patient's optimal joining-or-balking decision and whether or not to conduct doctor shopping. We find that doctor shopping is more likely to attain welfare improvement under the following conditions: the diagnosis accuracy is not high, the patients are more pessimistic and hold more diverse initial illness perceptions, and the relative value of identifying a severely ill patient is high. The subjective welfare maximization usually prefers a higher doctor shopping rate than the objective welfare maximization does. However, when the relative value of identifying a severely ill patient is high enough, the opposite holds true, which may explain why doctors usually encourage patients to conduct doctor shopping for serious health problems such as cancer.

Our work does not consider the incentive issues of the doctors' and assumes that both doctors and patients are risk-neutral. In practice, when misdiagnosis cost is very high and/or where people are highly risk-averse, repeated consultations and examinations may be recommended by doctors for the purpose of defensive medicine. For example, doctors in the United States are likely to encourage patients seek second opinions to reduce the risk of malpractice suits (see, e.g. King and Moulton (2006)), whereas doctors in mainland China often recommend patients to seek second opinions because of the highly risk-averse culture and/or for fear of violence against them (Zhang and Sleeboom-Faulkner, 2011). It would be interesting to study doctor shopping with risk-averse doctors and patients. Here, we consider the doctor-shopping behavior in a public diagnostic system where the price and service quality are both homogeneous across different service stations. It would be another interesting research to expand our work into a two-tier health care system with both for-profit and not-for-profit service facilities, and study how different levels of service quality and prices affect patients' doctor shopping decisions. Besides, online doctor shopping is an emerging phenomenon in the current era of digital economy, under which a patient can decide whether to consult multiple physicians *upfront* before the diagnostic result is released to her. Such online doctor shopping is likely a revenue maximization problem of the online platform. We would like to leave it for future research.

Acknowledgments

We are grateful to the departmental editor (Professor Sergei Savin), an anonymous senior editor, and two anonymous referees for their very helpful comments and suggestions. Dr. Fengfeng Huang's work was supported by the National Natural Science Foundation of China (Grant number: 72101048, 72032001). Dr. Pengfei Guo was supported in part by the Research Grants Council of Hong Kong (Grant number: 15506417). The corresponding author, Dr Yulan Wang is also affiliated with the Hong Kong Polytechnic University Shenzhen Research Institute and acknowledges the financial supports from the National Natural Science Foundation of China (Grant No. 71971184). All authors contributed equally to the work.

References

Alizamir, S., De Vericourt, F., Sun, P., 2013. Diagnosis accuracy under congestion. *Management Science*, 59(1), 157-171.

Althabe, F., Belizán, J.M., Villar, J., Alexander, S., Bergel, E., Ramos, S., Romero, M., Donner, A., Lindmark, G., Langer, A. and Farnot, U., 2004. Mandatory second opinion to reduce rates of unnecessary caesarean sections in Latin America: a cluster randomised controlled trial. *The Lancet*, 363(9425), pp.1934-1940.

Armstrong, D., Fry, J. and Armstrong, P., 1991. Doctors' perceptions of pressure from patients for referral. *British Medical Journal*, 302(6786), 1186-1188.

Chan, C. W., Yom-Tov, G., Escobar, G., 2014. When to use speedup: An examination of service systems with returns. *Operations Research*, 62(2), 462-482.

Chang, H.R., Yang, M.C. and Chung, K.P., 2013. Can cancer patients seeking a second opinion get better care?. *The American journal of managed care*, 19(5), 380-387.

Cui, S., Su, X. and Veeraraghavan, S., 2019. A model of rational retrials in queues. *Operations Research*, 67(6), 1699-1718.

Dai, T., Akan, M. and Tayur, S., 2017. Imaging room and beyond: The underlying economics behind physicians' test-ordering behavior in outpatient services. *Manufacturing & Service Operations Management*, 19(1), 99-113.

De Vericourt, F., Zhou, Y. P., 2005. Managing response time in a call-routing problem with service failure. *Operations Research*, 53(6), 968-981.

Donkin, L., Ellis, C. J., Powell, R., Broadbent, E., Gamble, G., Petrie, K. J., 2006. Illness perceptions predict reassurance following a negative exercise stress testing result. *Psychology and Health*, 21(4), 421-430.

Garcia, D., Spruill, L.S., Irshad, A., Wood, J., Kepecs, D. and Klauber-DeMore, N., 2018. The value of a second opinion for breast cancer patients referred to a National Cancer Institute (NCI)-designated cancer center with a multidisciplinary breast tumor board. Annals of surgical oncology, 25(10), 2953-2957.

Green, L. V., Savin, S., Wang, B., 2006. Managing patient service in a diagnostic medical facility. *Operations Research*, 54(1), 11-25.

Green, L.V. and Savin, S., 2008. Reducing delays for medical appointments: A queueing approach. *Operations Research*, 56(6), 1526-1538.

Guo, P., Tang, C.S., Wang, Y. and Zhao, M., 2016. The impact of reimbursement policy on patient welfare, readmission rate and waiting time in a public healthcare system: Fee-for-service vs. Bundle Payment. UCLA Anderson School of Management, Working Paper.

Guo, P., Lindsey, R., Zhang, Z. G., 2014. On the Downs-Thomson paradox in a self-financing two-tier queuing system. *Manufacturing & Service Operations Management*, 16(2), 315-322.

Hassin, R. and Roet-Green, R., 2018. Cascade equilibrium strategies in a two-server queueing system with inspection cost. *European Journal of Operational Research*, 267(3), 1014-1026.

Horn, S.D., Sharkey, P.D. and Gassaway, J., 1996. Managed Care Outcomes Project: study design, baseline patient characteristics, and outcome measures. *American Journal of Managed Care*, 2(3), 237-247.

Kasteler, J., Kane, R. L., Olsen, D. M., Thetford, C., 1976. Issues underlying prevalence of "doctor-shopping" behavior. *Journal of Health and Social Behavior*, 328-339.

King, J.S. and Moulton, B.W., 2006. Rethinking informed consent: the case for shared medical decision-making. *American journal of law & medicine*, 32(4), 429-501.

Leventhal, H. and Cameron, L., 1987. Behavioral theories and the problem of compliance. *Patient education and counseling*, 10(2), 117-138.

Leventhal, H., Zimmerman, R. and Gutmann, M., 1984. Compliance: a self-regulate perspective. In W.D. Gentry (Ed.), *Handbook of behavioral medicine* (pp.369-436). New York: Guilford Press.

Lo, A., A. Hedley, G. Pei, S. Ong, L. Ho, R. Fielding, L. Daniel., 1994. Doctor-shopping in Hong Kong: implications for quality of care. *International Journal for Quality in Health Care*, 6(4), 371-381.

Macpherson, A., M. Kramer, F. Ducharme, H. Yang, F. Bélanger., 2001. Doctor shopping before and after a visit to a paediatric emergency department. *Paediatrics & Child Health*. 6(6), 341.

Manion, E., Cohen, M.B. and Weydert, J., 2008. Mandatory second opinion in surgical pathology referral material: clinical consequences of major disagreements. The American journal of surgical pathology, 32(5), 732-737.

Morrow, M., Jagsi, R., Alderman, A.K., Griggs, J.J., Hawley, S.T., Hamilton, A.S., Graff, J.J. and Katz, S.J., 2009. Surgeon recommendations and receipt of mastectomy for treatment of breast cancer. *JAMA*, 302(14), 1551-1556.

O'Donnell, C.A., 2000. Variation in GP referral rates: what can we learn from the literature?. *Family practice*, 17(6), 462-471.

Pac, M. F., Veeraraghavan, S., 2010. Strategic diagnosis and pricing in expert services. Working paper, Wharton School, University of Pennsylvania, Philadelphia.

Payne, V.L., Singh, H., Meyer, A.N., Levy, L., Harrison, D. and Graber, M.L., 2014, May. Patient-initiated second opinions: systematic review of characteristics and impact on diagnosis, treatment, and satisfaction. *In Mayo Clinic Proceedings*. 89(5), 687-696.

Petrie K. J., Weinman J., 2012. Patients Perceptions of Their Illness: The Dynamo of Volition in Health Care. *Current Directions in Psychological Science*, 21(1):60-65

Petrie, K. J., Jago, L. A., Devcich, D. A., 2007. The role of illness perceptions in patients with medical conditions. *Current Opinion in Psychiatry*, 20(2), 163-167.

Qian, Q., Guo, P., Lindsey, R., 2017. Comparison of Subsidy Schemes for Reducing Waiting Times in Healthcare Systems. *Production and Operations Management*, 26(11), 2033-2049.

Sansone, R.A. and Sansone, L.A., 2012. Doctor shopping: a phenomenon of many themes. *Innovations in clinical neuroscience*, 9(11-12), 42

Sato T, M. Takeichi, M. Shiraham, T. Fukui, J. Gude., 1995. Doctor shopping patients and users of alternative medicine among Japanese primary care patients. *General Hospital Psychiatry*, 17:115-25.

Shaked, M. and Shanthikumar, J.G., 2007. Stochastic orders. Springer Science & Business Media.

Shi, P., Helm, J. E., Sebastian Heese, H., Mitchell, A. M., 2020. An Operational Framework for the Adoption and Integration of New Diagnostic Tests. *Production and Operations Management*, forthcoming.

Shmueli, L., Davidovitch, N., Pliskin, J.S., Balicer, R.D., Hekselman, I. and Greenfield, G., 2017. Seeking a second medical opinion: composition, reasons and perceived outcomes in Israel. *Israel journal of health policy research*, 6(1), 67.

Virji, A. and Britten, N., 1991. A study of the relationship between patients' attitudes and doctors' prescribing. Family Practice, 8(4), 314-319.

Wang, X., Debo, L. G., Scheller-Wolf, A., Smith, S. F., 2010. Design and analysis of diagnostic service centers. *Management Science*, 56(11), 1873-1890.

Webb, S. and Lloyd, M., 1994. Prescribing and referral in general practice: a study of patients' expectations and doctors' actions. *British Journal of General Practice*, 44(381), 165-169.

Weinman, J., Petrie, K. J., 1997. Illness perceptions: a new paradigm for psychosomatics?. *Journal of Psychosomatic Research*, 42(2), 113-116.

Yang, L., Debo, L.G. and Gupta, V., 2019. Search among queues under quality differentiation. *Management Science*, 65(8), 3605-3623.

Yom-Tov, G. B., Mandelbaum, A., 2014. Erlang-R: A time-varying queue with reentrant customers, in support of healthcare staffing. *Manufacturing & Service Operations Management*, 16(2), 283-299.

Zhang, X. and Sleeboom-Faulkner, M., 2011. Tensions between medical professionals and patients in mainland China. *Healthcare Ethics,* 20, 458.

Online Appendix
"Modeling Patients' Illness Perception and Equilibrium Analysis of Their Doctor Shopping Behavior"

Appendix A Proofs

Recall that $q_{ij} := q(s = j|t = i)$. We have

$$\begin{aligned} q_{00} &= q(s=0|t=0) = 1 - q(s=1|t=0) = 1 - q_{01}, \\ q_{11} &= q(s=1|t=1) = 1 - q(s=0|t=1) = 1 - q_{10}. \end{aligned}$$

Therefore, $q_{00} = q_{11}$ is equivalent to $q_{01} = q_{01}$. Moreover, by (3) and (4), we have

$$\begin{aligned} & g_0(g_0(x)) = \frac{q_{10}^2 x_n}{q_{10}^2 x_n + q_{00}^2(1-x_n)};\ g_1(g_1(x)) = \frac{q_{11}^2 x_n}{q_{11}^2 x_n + q_{01}^2(1-x_n)}; \\ & g_1(g_0(x)) = g_0(g_1(x)) = \frac{q_{11}q_{10}x}{q_{11}q_{10}x + q_{00}q_{01}(1-x)} = x. \end{aligned} \tag{16}$$

For the purpose of simplification, for $j \in \{0, 1\}$ and any arbitrary function $h(\alpha)$, let $g_j \cdot h(\alpha) = g_j(h(\alpha))$, $(g_j)^2 \cdot h(\alpha) = g_j(g_j(h(\alpha)))$ and so on and so forth. Similarly, let $(g_1 \cdot g_0)^1(\alpha) = g_1(g_0(\alpha))$, and $(g_1 \cdot g_0)^2(\alpha) = g_1(g_0(g_1(g_0(\alpha))))$ and so on and so forth. For any $i \in \mathcal{N} = \{1, 2, 3...\}$, we can obtain from the above equation that

$$(g_1 \cdot g_0)^i(\alpha) = \alpha;\ (g_1)^j \cdot (g_1 \cdot g_0)^i(\alpha) = (g_1)^j(\alpha);\ (g_0)^j \cdot (g_1 \cdot g_0)^i(\alpha) = (g_0)^j(\alpha). \tag{17}$$

We will use the above notations a lot in the following analysis.
Proof of Lemma 1: Recall that a function $h(x)$ is convex if and only if it satisfies

$$h(x + \varepsilon) - h(x) \geq h'(x)\varepsilon, \tag{18}$$

where $\varepsilon > 0$ is an arbitrarily small number. Now let us consider a finite horizon dynamic programming with a fixed number of periods. As the number of periods goes to infinity, the finite-horizon DP converges to our current DP problem. Let $v_k(x_n)$ denote the value function when there are k remaining periods. Then,

$$v_0(x_n) = r(x_n) \text{ and } v(x_n) = \lim_{k\to\infty} v_k(x_n).$$

The corresponding optimality equation and value iteration recursion can be written respectively as follows:

$$\begin{aligned} v_k(x_n) &= \max\left\{r(x_n),\ E\left[v_{k-1}(X_{n+1})|x_n\right] - C_p\right\}; \\ E\left[v_{k-1}(X_{n+1})|x_n\right] &= P(s=0|x_n)v_{k-1}\left(g_0(x_n)\right) + P(s=1|x_n)v_{k-1}\left(g_1(x_n)\right). \end{aligned}$$

Define $\pi_k^-(x_n)$ and $\pi_k^+(x_n)$ as follows:

$$\pi_k^-(x_n) = P(s=0|x_n)v_k\left(g_0(x_n)\right) \text{ and } \pi_k^+(x_n) = P(s=1|x_n)v_k\left(g_1(x_n)\right).$$

Then,

$$E\left[v_{k-1}(X_{n+1})|x_n\right] = \pi_{k-1}^-(x_n) + \pi_{k-1}^+(x_n).$$

Recall that $g_0(x_n)$ is given in (4). For an arbitrary number ε, we have

$$\begin{aligned} g_0(x_n+\varepsilon) - g_0(x_n) &= \frac{q_{10}x_n}{q_{10}(x_n+\varepsilon) + q_{00}(1-x_n-\varepsilon)} - \frac{q_{10}x_n}{q_{10}x_n + q_{00}(1-x_n)} \\ &= \frac{q_{00}q_{10}\varepsilon}{[q_{10}(x_n+\varepsilon) + q_{00}(1-x_n-\varepsilon)][q_{10}x_n + q_{00}(1-x_n)]}. \end{aligned}$$

Therefore,

$$g_0'(x_n) = \lim_{\varepsilon\to 0}\frac{g_0(x_n+\varepsilon) - g_0(x_n)}{\varepsilon} = \frac{q_{00}q_{10}}{[q_{10}x_n + q_{00}(1-x_n)]^2}.$$

Recall that $P(s=0|x_n) = q_{00}(1-x_n) + q_{10}x_n$. We can further show that

$$P(s=0|x_n+\varepsilon)\left[g_0(x_n+\varepsilon) - g_0(x_n)\right] = \frac{q_{00}q_{10}\varepsilon}{q_{10}x_n + q_{00}(1-x_n)} = P(s=0|x_n)g_0'(x_n)\varepsilon. \quad (19)$$

Obviously, $P(s=0|x_n) = q_{00}(1-x_n) + q_{10}x_n$ is linear, and hence

$$P(s=0|x_n+\varepsilon) - P(s=0|x_n) = \frac{dP(s=0|x_n)}{dx_n}\varepsilon. \quad (20)$$

First, we show that for an arbitrarily small number $\varepsilon > 0$,

$$\begin{aligned} \pi_0^-(x_n+\varepsilon) - \pi_0^-(x_n) &= P(s=0|x_n+\varepsilon)v_0(g_0(x_n+\varepsilon)) - P(s=0|x_n)v_0(g_0(x_n)) \\ &\geq P(s=0|x_n+\varepsilon)\left\{v_0(g_0(x_n)) + v_0'(g_0(x_n))\left[g_0(x_n+\varepsilon) - g_0(x_n)\right]\right\} \\ &\quad -P(s=0|x_n)v_0(g_0(x_n)) \\ &= [P(s=0|x_n+\varepsilon) - P(s=0|x_n)]v_0(g_0(x_n)) \\ &\quad +P(s=0|x_n+\varepsilon)v_0'(g_0(x_n))\left[g_0(x_n+\varepsilon) - g_0(x_n)\right] \\ &= \frac{dP(s=0|x_n)}{dx_n}v_0(g_0(x_n))\varepsilon + P(s=0|x_n)v_0'(g_0(x_n))g_0'(x_n)\varepsilon \\ &= (\pi_0^-)'(x_n)\varepsilon, \end{aligned}$$

where "$\geq$" follows from (18) and that $v_0(x_n) = r(x_n)$ is convex (as shown by (7) in the manuscript, $r(x_n)$ first linearly decreases and then linearly increases, and thus is convex), and the third "=" follows from (19) and (20). By the condition in (18), $\pi_0^-(x_n)$ is convex in x_n.

Similarly, we can show that $\pi_0^+(x_n)$ is convex. Therefore, $E\left[v_0(X_{n+1})|x_n\right] = \pi_0^-(x_n)+\pi_0^+(x_n)$ is convex. Thus, $v_1(x_n) = \max\{r(x_n),\ E\left[v_0(X_{n+1})|x_n\right] - C_p\}$ is convex in x_n.

Next, suppose that $v_k(x_n)$ is convex. We can show that for an arbitrarily small $\varepsilon > 0$,

$$\begin{aligned}
\pi_k^-(x_n+\varepsilon) - \pi_k^-(x_n) &= P(s=0|x_n+\varepsilon)v_k(g_0(x_n+\varepsilon)) - P(s=0|x_n)v_k(g_0(x_n)) \\
&\geq P(s=0|x_n+\varepsilon)\left\{v_k(g_0(x_n)) + v_k'(g_0(x_n))\left[g_0(x_n+\varepsilon) - g_0(x_n)\right]\right\} \\
&\quad -P(s=0|x_n)v_k(g_0(x_n)) \\
&= \frac{dP(s=0|x_n)}{dx_n}v_k(g_0(x_n))\varepsilon + P(s=0|x_n)v_k'(g_0(x_n))g_0'(x_n)\varepsilon \\
&= (\pi_k^-)'(x_n)\varepsilon.
\end{aligned}$$

Thus, $\pi_k^-(x_n)$ is convex. Similarly, we can show that $\pi_k^+(x_n)$ is convex. Therefore, $E\left[v_k(X_{n+1})|x_n\right] = \pi_k^-(x_n) + \pi_k^+(x_n)$ is convex in x_n and thus,

$$v_{k+1}(x_n) = \max\{r(x_n),\ E\left[v_k(X_{n+1})|x_n\right] - C_p\}$$

is convex. By mathematical induction, $v(x_n) = \lim_{k\to\infty} v_k(x_n)$ is convex in x_n.

Derivation of the Condition That Ensures (9) **Holds:** It can be easily shown that

$$E\left[v(X_{n+1})|x_n\right] = P(s=0|x_n)v(g_0(x_n)) + P(s=1|x_n)v(g_1(x_n)). \tag{21}$$

By (1), (2), (3), and (4), we have

$$P(s=0|g_0(x_n)) = \frac{q_{10}^2x_n + q_{00}^2(1-x_n)}{q_{10}x_n + q_{00}(1-x_n)},\ P(s=1|g_0(x_n)) = \frac{q_{00}q_{01}(1-x_n) + q_{11}q_{10}x_n}{q_{10}x_n + q_{00}(1-x_n)}; \tag{22}$$

$$P(s=1|g_1(x_n)) = \frac{q_{01}^2(1-x_n) + q_{11}^2x_n}{q_{01}(1-x_n) + q_{11}x_n},\quad P(s=0|g_1(x_n)) = \frac{q_{00}q_{01}(1-x_n) + q_{11}q_{10}x_n}{q_{01}(1-x_n) + q_{11}x_n}. \tag{23}$$

By utilizing (16), (21), (22) and the convexity of $v(x_n)$ (see Lemma 1), we can show that

$$\begin{aligned}
E\left[v(X_{n+2})|g_0(x_n)\right] &= P(s=0|g_0(x_n))v(g_0\cdot g_0(x_n)) + P(s=1|g_0(x_n))v(g_1\cdot g_0(x_n)) \\
&= P(s=0|g_0(x_n))v(g_0\cdot g_0(x_n)) + P(s=1|g_0(x_n))v(x_n) \\
&\leq P(s=0|g_0(x_n))\left[(1-g_0\cdot g_0(x_n))v(0) + g_0\cdot g_0(x_n)v(1)\right] \\
&\quad +P(s=1|g_0(x_n))\left[(1-x_n)v(0) + x_nv(1)\right] \\
&= (1-g_0(x_n))v(0) + g_0(x_n)v(1).
\end{aligned}$$

Similarly, by utilizing (16), (21), (23) and the convexity of $v(x_n)$, we can show that

$$\begin{aligned}
E\left[v(X_{n+2})|g_1(x_n)\right] &= P(s=0|g_1(x_n))v(g_0\cdot g_1(x_n)) + P(s=1|g_1(x_n))v(g_1\cdot g_1(x_n)) \\
&= P(s=0|g_1(x_n))v(x_n) + P(s=1|g_1(x_n))v(g_1\cdot g_1(x_n)) \\
&\leq P(s=0|g_1(x_n))\left[(1-x_n)v(0) + x_nv(1)\right] \\
&\quad +P(s=1|g_1(x_n))\left[(1-g_1\cdot g_1(x_n))v(0) + g_1\cdot g_1(x_n)v(1)\right] \\
&= (1-g_1(x_n))v(0) + g_1(x_n)v(1).
\end{aligned}$$

By using (7) and the boundary conditions $v(0) = V_0$ (an illness perception $x_n = 0$ indicates that the patient is surely healthy and thus has a reward V_0) and $v(1) = V_1$ (an illness perception $x_n = 1$ indicates that the patient is surely ill and thus has a reward V_1), we then have

$$\begin{cases} E\left[v(X_{n+2})|g_0(x_n)\right] - C_p - r(g_0(x_n)) = g_0(x_n)(V_1 + L_1) - C_p & \text{if } 0 < x_n < \widehat{\alpha}. \\ E\left[v(X_{n+2})|g_1(x_n)\right] - C_p - r(g_1(x_n)) = [1 - g_1(x_n)](V_0 + L_0) - C_p & \text{otherwise.} \end{cases} \tag{24}$$

The optimality equation (6) indicates that a patient shall stop visiting/revisiting the diagnosis system if and only if

$$E\left[v(X_{n+1})|x_n\right] - C_p - r(x_n) \le 0.$$

Then, to ensure that the requirement

$$\begin{cases} g_0(x_n) \in S, & \text{if } 0 < x_n < \widehat{\alpha}; \\ g_1(x_n) \in S, & \text{otherwise} \end{cases} \tag{25}$$

holds, based on (24), it is equivalent to require that

$$\begin{cases} g_0(x_n)(V_1 + L_1) - C_p < 0, & \text{if } 0 < x_n < \widehat{\alpha}; \\ [1 - g_1(x_n)](V_0 + L_0) - C_p < 0, & \text{otherwise} \end{cases} \tag{26}$$

From (3) and (4), we can obtain that $g_0(x_n)$ and $g_1(x_n)$ increase in x_n and

$$g_1(x_n) - x_n = \frac{(q_{11} - q_{01})x_n(1 - x_n)}{q_{11}x_n + q_{01}(1 - x_n)} > 0, \;\; g_0(x_n) - x_n = \frac{(q_{10} - q_{00})x_n(1 - x_n)}{q_{10}x_n + q_{00}(1 - x_n)} < 0.$$

Then, (26) holds if

$$C_p \ge \max\left\{g_0(\widehat{\alpha})(V_1 + L_1),\;\; [1 - g_1(\widehat{\alpha})](V_0 + L_0)\right\},$$

under which

$$\begin{cases} v(g_0(x_n)) = r(g_0(x_n)) \ge E\left[v(X_{n+2})|g_0(x_n)\right] - C_p & \text{if } 0 < x_n < \widehat{\alpha} \\ v(g_1(x_n)) = r(g_1(x_n)) \ge E\left[v(X_{n+2})|g_1(x_n)\right] - C_p & \text{otherwise} \end{cases}. \tag{27}$$

For the patient with an illness perception $\widehat{\alpha} = \frac{V_0+L_0}{V_0+V_1+L_0+L_1}$, to ensure that $\widehat{\alpha} \notin S$, we need

$$v(\widehat{\alpha}) = E\left[v(X_{n+1})|\widehat{\alpha}\right] - C_p > r(\widehat{\alpha}),$$

By utilizing (1), (2), (3), (4), (7), (8), (21) and (27), we can show that it satisfies

$$\begin{aligned} E\left[v(X_{n+1})|\widehat{\alpha}\right] - C_p - r(\widehat{\alpha}) &= P(s = 0|\widehat{\alpha})v(g_0(\widehat{\alpha})) + P(s = 1|\widehat{\alpha})v(g_1(\widehat{\alpha})) - r(\widehat{\alpha}) - C_p \\ &= P(s = 0|\widehat{\alpha})r(g_0(\widehat{\alpha})) + P(s = 1|\widehat{\alpha})r(g_1(\widehat{\alpha})) - r(\widehat{\alpha}) - C_p \\ &= (q_{11} - q_{01})\widehat{\alpha}(1 - \widehat{\alpha})(V_0 + V_1 + L_0 + L_1) - C_p > 0 \end{aligned} \tag{28}$$

whenever $C_p < (q_{11} - q_{10})\widehat{\alpha}(1-\widehat{\alpha})(V_0 + V_1 + L_0 + L_1)$. Based on the above analysis, we can show that (9) always holds if

$$\max\{g_0(\widehat{\alpha})(1-\widehat{\alpha}),\ [1-g_1(\widehat{\alpha})]\widehat{\alpha}\} \leq \frac{C_p}{V_0+V_1+L_0+L_1} < (q_{11}-q_{10})\widehat{\alpha}(1-\widehat{\alpha}).$$

Proof of Proposition 1: A patient shall visit or continue to visit the system if and only if

$$E\left[v(X_{n+1})|x_n\right] - C_p - r(x_n) > 0. \tag{29}$$

Recall that $E\left[v(X_{n+1})|x_n\right]$ is convex as shown in the proof of Lemma 1, and $r(x_n)$ stated in (7) is linear over the two intervals $0 < x_n < \widehat{\alpha}$ and $\widehat{\alpha} \leq x_n < 1$, respectively. Consider a special case of $x_n = 0 < \widehat{\alpha}$. We can show that

$$\begin{aligned}
E\left[v(X_{n+1})|0\right] - C_p - r(0) &= P(s=0|0)v(g_0(0)) + P(s=1|0)v(g_1(0)) - C_p - r(0)\\
&= P(s=0|0)r(g_0(0)) + P(s=1|0)r(g_1(0)) - C_p - r(0)\\
&= -C_p < 0,
\end{aligned}$$

where the first "=" follows from (18), the second "=" follows from (24), and the third "=" follows from $g_1(0) = 0$ and $g_0(0) = 0$ ($g_1(x_n)$ and $g_0(x_n)$ are given by (3) and (4), respectively). Similarly, we can show

$$E\left[v(X_{n+1})|1\right] - C_p - r(1) = -C_p < 0.$$

Recall that in (28), we have $E\left[v(X_{n+1})|\widehat{\alpha}\right] - C_p - r(\widehat{\alpha}) > 0$. Therefore, as x_n increases, $r(x_n)$ crosses $E\left[v(X_{n+1})|x_n\right] - C_p$ exactly once from above in the interval $0 < x_n < \widehat{\alpha}$. We denote the intersection point as $\underline{\alpha}$. Similarly, $r(x_n)$ crosses $E\left[v(X_{n+1})|x_n\right] - C_p$ exactly once from below in the interval $\widehat{\alpha} \leq x_n < 1$ at a point denoted by $\overline{\alpha}$. That is,

$$E\left[v(X_{n+1})|x_n\right] - C_p - r(x_n) \leq 0, \text{ when } x_n \in (0, \underline{\alpha}] \cup [\overline{\alpha}, 1),$$

under which it is optimal for the patient not to visit/revisit the system.

Note that when the patient's initial illness perception is above the upper threshold, namely when $\overline{\alpha} \leq x_0 = \alpha < 1$, she still needs to join the system because diagnosis is required prior to treatment. As such, the optimal stopping set S that characterizes the set of illness perceptions with which it is optimal for the patient not to visit/revisit the system can be written as follows:

$$S = \{X_n | 0 < X_n \leq \underline{\alpha} \text{ or } \overline{\alpha} \leq X_{n+1} < 1,\ n = 0, 1, 2, ...\}.$$

Since both $g_0(x_n)$ and $g_1(x_n)$ increase in x_n, (25) holds whenever

$$g_0(\widehat{\alpha}) \leq \underline{\alpha} \text{ and } g_1(\widehat{\alpha}) \geq \overline{\alpha}. \tag{30}$$

By using (16), we show $\overline{\alpha} \leq g_1(\widehat{\alpha}) = g_1 \cdot g_1 \cdot g_0(\widehat{\alpha}) \leq g_1 \cdot g_1(\underline{\alpha})$ and $\underline{\alpha} \geq g_0(\widehat{\alpha}) = g_0 \cdot g_0 \cdot g_1(\widehat{\alpha}) \geq g_0 \cdot g_0(\overline{\alpha})$, which is

$$\begin{aligned} \underline{\alpha} - g_0(g_0(\overline{\alpha})) &= \frac{q_{00}^2 \overline{\alpha}(1-\underline{\alpha})}{q_{10}^2 \overline{\alpha} + q_{00}^2(1-\overline{\alpha})} \left(\frac{\underline{\alpha}(1-\overline{\alpha})}{\overline{\alpha}(1-\underline{\alpha})} - \frac{q_{10}^2}{q_{00}^2} \right) \geq 0, \\ g_1(g_1(\underline{\alpha})) - \overline{\alpha} &= \frac{q_{11}^2 \overline{\alpha}(1-\underline{\alpha})}{q_{11}^2 \underline{\alpha} + q_{01}^2(1-\underline{\alpha})} \left(\frac{\underline{\alpha}(1-\overline{\alpha})}{\overline{\alpha}(1-\underline{\alpha})} - \frac{q_{01}^2}{q_{11}^2} \right) \geq 0. \end{aligned} \tag{31}$$

Since $g_1(x)$ and $g_0(x)$ both increase in x, we can show that for $x_n \in (\underline{\alpha},\ \overline{\alpha})$,

$$g_1 \cdot g_1(x_n) > g_1 \cdot g_1(\underline{\alpha}) \geq \overline{\alpha} \text{ and } g_0 \cdot g_0(x_n) < g_0 \cdot g_0(\overline{\alpha}) \leq \underline{\alpha}; \tag{32}$$

that is, whenever $x_n \in (\underline{\alpha},\ \overline{\alpha})$, the patient leaves the system after obtaining the same diagnosis results at two successive visits.

We now derive the expressions of the two thresholds $\underline{\alpha}$ and $\overline{\alpha}$. First, consider a patient with an illness perception of $\underline{\alpha} < x_n < \widehat{\alpha}$. Since $g_0(x_n)$ increases in x_n, we show from (7) and (30) that, respectively,

$$r(x_n) = (1-x_n)V_0 - x_n L_1 \text{ and } g_0(x_n) < g_0(\widehat{\alpha}) \leq \underline{\alpha}. \tag{33}$$

We shall consider the following two scenarios:

1a. $g_1(x_n) \geq \overline{\alpha}$, namely, $g_1(x_n) \in S$;

1b. $g_1(x_n) < \overline{\alpha}$, namely, $g_1(x_n) \notin S$;

We examine the above scenarios as follows.

Scenario 1a: $\boldsymbol{g_1(x_n) \in S}$. Obviously, $v(g_0(x_n)) = r(g_0(x_n))$ and $v(g_1(x_n)) = r(g_1(x_n))$; from (21), we show

$$E\left[v(X_{n+1})|x_n\right] = P(s=0|x_n) r(g_0(x_n)) + P(s=1|x_n) r(g_1(x_n)) = E\left[r(X_{n+1})|x_n\right],$$

which indicates the One-Step Look-Ahead (OSLA) rule is optimal under Scenario 1a. By using (1), (2), (4), (3), (7), and the continuing condition (29), we show that the patient shall continue if and only if

$$E\left[v(X_{n+1})|x_n\right] - C_p - r(x_n) = q_{11} x_n (V_1 + L_1) - q_{01}(1-x_n)(V_0 + L_0) - C_p > 0.$$

We then show that the patient shall continue if and only if

$$x_n > \frac{q_{01}(L_0 + V_0) + C_p}{q_{11}(L_1 + V_1) + q_{01}(L_0 + V_0)} := \underline{s}^{1a}.$$

Scenario 1b: $g_1(x_n) \notin S$. Under this scenario, we have

$$\begin{aligned} v(g_0(x_n)) &= r(g_0(x_n)) = \frac{q_{00}(1-x_n)V_0 - q_{10}x_nL_1}{q_{00}(1-x_n)+q_{10}x_n}; \qquad (34) \\ v(g_1(x_n)) &= \max\{r(g_1(x_n)),\ E\left[v(X_{n+2})|g_1(x_n)\right] - C_p\} = E\left[v(X_{n+2})|g_1(x_n)\right] - C_p \\ &= P(s=1|g_1(x_n))v(g_1\cdot g_1(x_n)) + P(s=0|g_1(x_n))v(g_0\cdot g_1(x_n)) - C_p \\ &= P(s=1|g_1(x_n))r(g_1\cdot g_1(x_n)) + P(s=0|g_1(x_n))v(x_n) - C_p; \\ v(x_n) &= P(s=1|x_n)v(g_1(x_n)) + P(s=0|x_n)v(g_0(x_n)) - C_p, \end{aligned}$$

where the first line follows from (4), $g_0(x_n) < x_n < \widehat{\alpha}$, and (33, the second line follows from $g_1(x_n) \notin S$ and (6), the fourth line follows from (17) and (32), and the last line follows from $x_n \notin S$ and (6). By plugging the expression of $v(x_n)$ into $v(g_1(x_n))$, we can derive that

$$\begin{aligned} v(g_1(x_n)) &= \frac{P(s=1|g_1(x_n))r(g_1\cdot g_1(x_n)) + P(s=0|x_n)P(s=0|g_1(x_n))v(g_0(x_n))}{1-P(s=1|x_n)P(s=0|g_1(x_n))} \\ &\quad -\frac{P(s=0|g_1(x_n))+1}{1-P(s=1|x_n)P(s=0|g_1(x_n))}C_p. \qquad (35) \end{aligned}$$

Recall that $P(s=1|g_1(x_n))$ and $P(s=0|g_1(x_n))$ are both given in (23). From (30), we can show $g_1\cdot g_1(x_n) > \overline{\alpha} > \widehat{\alpha}$, and moreover, by (16) and (7), we have

$$r(g_1\cdot g_1(x_n)) = g_1\cdot g_1(x_n)V_1 - (1-g_1\cdot g_1(x_n))L_0 = \frac{q_{11}^2x_nV_1 - q_{01}^2(1-x_n)L_0}{q_{01}^2(1-x_n)+q_{11}^2x_n}.$$

Plugging the above equations and (34) into (35), we can get

$$\begin{aligned} v(g_1(x_n)) &= \frac{q_{11}^2x_nV_1 - q_{01}^2(1-x_n)L_0 + [q_{00}q_{01}(1-x_n)+q_{11}q_{10}x_n][q_{00}(1-x_n)V_0 - q_{10}x_nL_1]}{[q_{01}(1-x_n)+q_{11}x_n][1-(q_{00}q_{01}(1-x_n)+q_{11}q_{10}x_n)]} \\ &\quad - \frac{q_{01}(1-x_n)+q_{11}x_n+q_{00}q_{01}(1-x_n)+q_{11}q_{10}x_n}{[q_{01}(1-x_n)+q_{11}x_n][1-(q_{00}q_{01}(1-x_n)+q_{11}q_{10}x_n)]}C_p \\ &= \frac{q_{11}^2x_nV_1 - q_{01}^2(1-x_n)L_0 + q_{00}q_{01}[q_{00}(1-x_n)V_0 - q_{10}x_nL_1]}{[q_{01}(1-x_n)+q_{11}x_n](1-q_{00}q_{01})} \\ &\quad - \frac{q_{01}(1-x_n)+q_{11}x_n+q_{00}q_{01}}{[q_{01}(1-x_n)+q_{11}x_n](1-q_{00}q_{01})}C_p. \qquad (36) \end{aligned}$$

Recall that the patient continues only if

$$E[v(X_{n+1})|x_n] - r(x_n) - C_p = P(s=0|x_n)r(g_0(x_n)) + P(s=1|x_n)v(g_1(x_n)) - r(x_n) - C_p > 0.$$

Using (1), (2), (33), (34), and (36), we can show that it requires

$$x_n > \frac{q_{01}^2(L_0+V_0) + (1+q_{01})C_p}{q_{01}^2(L_0+V_0) + q_{11}^2(L_1+V_1) - (q_{11}-q_{01})C_p} := \underline{s}^{1b}.$$

Obviously, $\underline{\alpha} = \min\{\underline{s}^{1a},\ \underline{s}^{1b}\}$, which is

$$\underline{\alpha} \;=\; \begin{cases} \underline{s}^{1b} := \frac{q_{01}^2(L_0+V_0)+(1+q_{01})C_p}{q_{01}^2(L_0+V_0)+q_{11}^2(L_1+V_1)-(q_{11}-q_{01})C_p} & \text{if } 0 < C_p < \mathcal{C}, \\ \underline{s}^{1a} := \frac{q_{01}(L_0+V_0)+C_p}{q_{11}(L_1+V_1)+q_{01}(L_0+V_0)} & \text{otherwise;} \end{cases}$$

Next, we look into the interval of $\widehat{\alpha} < \overline{s} \le x_n < \overline{\alpha}$. Since $g_1(x_n)$ increases in x_n, from (7) and (30), we show, respectively,

$$r(x_n) = x_n V_1 - (1-x_n)L_0 \text{ and } g_1(x_n) > g_1(\widehat{\alpha}) \ge \overline{\alpha}. \tag{37}$$

there exist the following two scenarios:

2a. $g_0(x_n) \le \underline{\alpha}$, namely, $g_0(x_n) \in S$;

2b. $g_0(x_n) > \underline{\alpha}$, namely, $g_0(x_n) \notin S$;

Scenario 2a: $\boldsymbol{g_0(x_n) \in S}$. Here, $v(g_0(x_n)) = r(g_0(x_n))$ and $v(g_1(x_n)) = r(g_1(x_n))$. Similar to Scenario 1a, the One-Step Look-Ahead (OSLA) rule is optimal under Scenario 2a. From (21) and (29), we show that the patient shall continue if and only if

$$\begin{aligned} E\left[v(X_{n+1})|x_n\right] - r(x_n) - C_p &= P(s=1|x_n)r(g_1(x_n)) + P(s=0|x_n)r(g_0(x_n)) - r(x_n) - C_p \\ &= q_{00}(1-x_n)(V_0+L_0) - q_{10}x_n(V_1+L_1) - C_p > 0, \end{aligned}$$

where the second "=" follow from (1), (2), (4), (3), and (7). We then show that the patient shall continue if and only if

$$x_n < \frac{q_{00}(L_0+V_0) - C_p}{q_{00}(L_0+V_0) + q_{10}(L_1+V_1)} := \overline{s}^{2a}.$$

Scenario 2b: $\boldsymbol{g_0(x_n) \notin S}$. Under this scenario, we have

$$\begin{aligned} v(g_1(x_n)) &= r(g_1(x_n)) = \frac{q_{11}x_nV_1 - q_{01}(1-x_n)L_0}{q_{11}x_n + q_{01}(1-x_n)}; \qquad (38) \\ v(g_0(x_n)) &= \max\{r(g_0(x_n)),\ E\left[v(X_{n+2})|g_0(x_n)\right] - C_p\} = E\left[v(X_{n+2})|g_0(x_n)\right] - C_p \\ &= P(s=1|g_0(x_n))v(g_1\cdot g_0(x_n)) + P(s=0|g_0(x_n))v(g_0\cdot g_0(x_n)) - C_p; \\ &= P(s=1|g_0(x_n))v(x_n) + P(s=0|g_0(x_n))r(g_0\cdot g_0(x_n)) - C_p, \\ v(x_n) &= P(s=0|x_n)v(g_0(x_n)) + P(s=1|x_n)v(g_1(x_n)) - C_p. \end{aligned}$$

where the first line follows from (3), $g_0(x_n) < x_n < \widehat{\alpha}$, and (33, the second line follows from $g_0(x_n) \notin S$ and (6), the fourth line follows from (17) and (32), and the last line follows from $x_n \notin S$ and (6). By plugging the expression of $v(x_n)$ into $v(g_0(x_n))$, we can derive

$$\begin{aligned} v(g_0(x_n)) &= \frac{P(s=0|g_0(x_n))r(g_0\cdot g_0(x_n)) + P(s=1|x_n)P(s=1|g_0(x_n))v(g_1(x_n))}{1 - P(s=0|x_n)P(s=1|g_0(x_n))} \\ &\quad - \frac{P(s=1|g_0(x_n)) + 1}{1 - P(s=0|x_n)P(s=1|g_0(x_n))}C_p. \end{aligned} \tag{39}$$

From (30), we show $g_0 \cdot g_0(x_n) < \underline{\alpha} < \widehat{\alpha}$, and moreover, by (4) and (7), we have

$$r(g_0 \cdot g_0(x_n)) = (1 - g_0 \cdot g_0(x_n))V_0 - g_0 \cdot g_0(x_n)L_1 = \frac{q_{00}^2(1-x_n)V_0 - q_{10}^2 x_n L_1}{q_{10}^2 x_n + q_{00}^2(1-x_n)}.$$

Recall that $P(s=0|g_0(x_n))$ and $P(s=1|g_0(x_n))$ are given in (22). Plugging the above equations and (38) into (39), we can get

$$\begin{aligned} v(g_0(x_n)) &= \frac{q_{00}^2(1-x_n)V_0 - q_{10}^2 x_n L_1 + [q_{00}q_{01}(1-x_n) + q_{11}q_{10}x_n][q_{11}x_n V_1 - q_{01}(1-x_n)L_0]}{(q_{10}x_n + q_{00}(1-x_n))[1-(q_{00}q_{01}(1-x_n)+q_{11}q_{10}x_n)]} \\ &\quad - \frac{q_{10}x_n + q_{00}(1-x_n) + q_{00}q_{01}(1-x_n) + q_{11}q_{10}x_n}{(q_{10}x_n + q_{00}(1-x_n))[1-(q_{00}q_{01}(1-x_n)+q_{11}q_{10}x_n)]} C_p \\ &= \frac{q_{00}^2(1-x_n)V_0 - q_{10}^2 x_n L_1 + q_{00}q_{01}[q_{11}x_n V_1 - q_{01}(1-x_n)L_0]}{(q_{10}x_n + q_{00}(1-x_n))(1-q_{00}q_{01})} \\ &\quad - \frac{q_{10}x_n + q_{00}(1-x_n) + q_{00}q_{01}}{(q_{10}x_n + q_{00}(1-x_n))(1-q_{00}q_{01})} C_p. \end{aligned} \tag{40}$$

Recall that the patient continues only if

$$E[v(X_{n+1})|x_n] - C_p - r(x_n) = P(s=0|x_n)v(g_0(x_n)) + P(s=1|x_n)r(g_1(x_n)) - r(x_n) - C_p > 0.$$

In a similar vein, following that in Scenario 1c, we can show that $P(s=1|x_n)$, $P(s=0|x_n)$, $r(g_1(x_n))$ and $v(g_0(x_n))$ in the above continuing condition are given by (1), (2), (38), and (40), respectively, except that here, $r(x_n)$ is given by (37) while in Scenario 1c, $r(x_n)$ is given by (33). We then can show that the continuing condition requires

$$x_n < \frac{q_{00}^2(L_0+V_0) - (1+q_{00})C_p}{q_{00}^2(L_0+V_0) + q_{10}^2(L_1+V_1) - (q_{00}-q_{10})C_p} := \overline{s}^{2b}.$$

Obviously, $\overline{\alpha} = \max\{\overline{s}^{2a}, \ \overline{s}^{2b}\}$, which is,

$$\overline{\alpha} = \begin{cases} \overline{s}^{2b} := \frac{q_{00}^2(L_0+V_0)-(1+q_{00})C_p}{q_{00}^2(L_0+V_0)+q_{10}^2(L_1+V_1)-(q_{00}-q_{10})C_p} & \text{if } 0 < C_p < \mathcal{C}, \\ \overline{s}^{2a} := \frac{q_{00}(L_0+V_0)-C_p}{q_{00}(L_0+V_0)+q_{10}(L_1+V_1)} & \text{otherwise.} \end{cases}$$

Proof of Lemma 2: From (31), we can see that (9) also requires $\eta > \left(\frac{q_{10}}{q_{11}}\right)^2$. From (10) and (11), we have

$$\frac{d\underline{\alpha}}{dC_p} = \begin{cases} \frac{q_{01}^2(1+q_{11})(L_0+V_0)+q_{11}^2(1+q_{01})(L_1+V_1)}{\left[q_{01}^2(L_0+V_0)+q_{11}^2(L_1+V_1)-(q_{11}-q_{01})C_p\right]^2} > 0 & \text{if } 0 < C_p < \mathcal{C} \\ \frac{1}{q_{11}(L_1+V_1)+q_{01}(L_0+V_0)} > 0, & \text{otherwise.} \end{cases} \tag{41}$$

$$\frac{d\overline{\alpha}}{dC_p} = \begin{cases} -\frac{q_{00}^2(1+q_{10})(L_0+V_0)+q_{10}^2(1+q_{00})(L_1+V_1)}{\left[q_{00}^2(L_0+V_0)+q_{10}^2(L_1+V_1)-(q_{00}-q_{10})C_p\right]^2} < 0 & \text{if } 0 < C_p < \mathcal{C} \\ -\frac{1}{q_{00}(L_0+V_0)+q_{10}(L_1+V_1)} < 0, & \text{otherwise.} \end{cases} \tag{42}$$

Obviously, $\eta := \frac{\underline{\alpha}(1-\overline{\alpha})}{\overline{\alpha}(1-\underline{\alpha})}$ increases in $\underline{\alpha}$ and decreases in $\overline{\alpha}$, and hence, η increases in C_p. From (31), we show that $\eta > \left(\frac{q_{10}}{q_{11}}\right)^2$. Since $\underline{\alpha} < \overline{\alpha}$, we show that $\eta < 1$. It can be easily shown that when $C_p = \mathcal{C}$,

$$\eta = \frac{(q_{01}\widehat{\alpha} + \theta)(q_{10}(1-\widehat{\alpha}) + \theta)}{(q_{00}\widehat{\alpha} - \theta)(q_{11}(1-\widehat{\alpha}) - \theta)} = \frac{q_{10}}{q_{11}}.$$

Therefore,

$$\begin{cases} \left(\frac{q_{10}}{q_{11}}\right)^2 \le \eta < \frac{q_{10}}{q_{11}} & \text{if } \ 0 < C_p < \mathcal{C}, \\ \frac{q_{10}}{q_{11}} \le \eta < 1, & \text{otherwise.} \end{cases}$$

Recall that

$$\underline{\alpha}_b = \frac{\overline{\alpha} q_{01}}{\overline{\alpha} q_{01} + (1-\overline{\alpha}) q_{11}}, \tag{43}$$

$$\overline{\alpha}_b = \frac{\underline{\alpha} q_{00}}{\underline{\alpha} q_{00} + (1-\underline{\alpha}) q_{10}}. \tag{44}$$

$$\alpha_s = \frac{\overline{\alpha} q_{00}}{\overline{\alpha} q_{00} + (1-\overline{\alpha}) q_{10}}. \tag{45}$$

We show the following:

$$\begin{aligned}
\underline{\alpha}_b - \overline{\alpha}_b &= \frac{\overline{\alpha}(1-\underline{\alpha}) q_{00} q_{11}}{[\overline{\alpha} q_{01} + (1-\overline{\alpha}) q_{11}][\underline{\alpha} q_{00} + (1-\underline{\alpha}) q_{10}]} \left(\left(\frac{q_{10}}{q_{11}}\right)^2 - \eta \right) < 0, \\
\underline{\alpha}_b - \underline{\alpha} &= \frac{\overline{\alpha}(1-\underline{\alpha}) q_{11}}{\overline{\alpha} q_{01} + (1-\overline{\alpha}) q_{11}} \left(\frac{q_{10}}{q_{11}} - \eta \right), \\
\overline{\alpha} - \overline{\alpha}_b &= \frac{\overline{\alpha}(1-\underline{\alpha}) q_{00}}{\underline{\alpha} q_{00} + (1-\underline{\alpha}) q_{10}} \left(\frac{q_{10}}{q_{00}} - \eta \right), \\
\overline{\alpha}_b - \alpha_s &= \frac{\overline{\alpha}(1-\underline{\alpha}) q_{00} q_{10}}{[\underline{\alpha} q_{00} + (1-\underline{\alpha}) q_{10}][\overline{\alpha} q_{00} + (1-\overline{\alpha}) q_{10}]} (\eta - 1) < 0,
\end{aligned}$$

When $\left(\frac{q_{10}}{q_{11}}\right)^2 \le \eta < \frac{q_{10}}{q_{11}}$, $\underline{\alpha} < \underline{\alpha}_b \le \overline{\alpha}_b < \overline{\alpha} < \alpha_s$ and when $\frac{q_{01}}{q_{11}} \le \eta < 1$, $\underline{\alpha}_b \le \underline{\alpha} < \overline{\alpha} \le \overline{\alpha}_b < \alpha_s$.

Proof of Proposition 2: It can be shown from (4), (4), (43), (44), and (45), that $g_1(\underline{\alpha}_b) = \overline{\alpha}$, $g_0(\overline{\alpha}_b) = \underline{\alpha}$, and $g_0(\alpha_s) = \overline{\alpha}$.

When $\left(\frac{q_{10}}{q_{11}}\right)^2 \le \eta < \frac{q_{10}}{q_{11}}$, we can show that if $\underline{\alpha} < \alpha < \underline{\alpha}_b$, $g_1(\alpha) < g_1(\underline{\alpha}_b) = \overline{\alpha}$ and $g_0(\alpha) < g_0(\overline{\alpha}_b) = \underline{\alpha}$; hence, the patient leaves the system only if she obtains a negative result. If $\underline{\alpha}_b \le \alpha \le \overline{\alpha}_b$, $g_1(\alpha) \ge g_1(\underline{\alpha}_b) = \overline{\alpha}$ and $g_0(\alpha) \le g_0(\overline{\alpha}_b) = \underline{\alpha}$; hence, the patient leaves the system regardless of the result; If $\overline{\alpha}_b < \alpha < \alpha_s$, $g_1(\alpha) > g_1(\underline{\alpha}_b) = \overline{\alpha}$ and $g_0(\alpha) > g_0(\overline{\alpha}_b) = \underline{\alpha}$; hence, the patient leaves the system only if she obtains a positive result; If $\alpha_s \le \alpha < 1$, both $g_1(\alpha) > \alpha > \overline{\alpha}$ and $g_0(\alpha) \geq g_0(\alpha_s) \geq \overline{\alpha}$; hence, the patient leaves the system regardless of the diagnosis result. Moreover, since her updated illness perception is always above the upper threshold $\overline{\alpha}$, she presses for treatment referral.

When $\frac{q_{01}}{q_{11}} \leq \eta < 1$, we can show that if $\underline{\alpha} \leq \alpha \leq \overline{\alpha}_b$, $g_1(\alpha) \geq g_1(\underline{\alpha}_b) = \overline{\alpha}$ and $g_0(\alpha) \leq g_0(\overline{\alpha}_b) = \underline{\alpha}$; hence, the patient leaves the system regardless of the result. For $\overline{\alpha}_b < \alpha \leq \alpha_s$ and $\alpha_s \leq \alpha < 1$, it is the same as that when $\left(\frac{q_{10}}{q_{11}}\right)^2 \leq \eta < \frac{q_{10}}{q_{11}}$.

Proof of Proposition 3: It is easy to show that when $0 < y < 1$,

$$\sum_{i=0}^{\infty} iy^i = \frac{y}{(1-y)^2}. \tag{46}$$

Since the obedient and stubborn patients always visit once. The following analysis focus exclusively on the diagnosis-dependent patient.

Case 1: $\left(\frac{q_{01}}{q_{11}}\right)^2 \leq \eta < \frac{q_{01}}{q_{11}}$. Recall that

$$\widetilde{\alpha} = \frac{\underline{\alpha} q_{00}^2}{\underline{\alpha} q_{00}^2 + (1-\underline{\alpha}) q_{10}^2}. \tag{47}$$

It can be easily shown that $g_0(\widetilde{\alpha}) := \overline{\alpha}_b$; see (4) and (11). From (11) and (45), we show

$$\begin{aligned} \widetilde{\alpha} - \overline{\alpha} &= \frac{\overline{\alpha}(1-\underline{\alpha})q_{00}^2}{\underline{\alpha} q_{00}^2 + (1-\underline{\alpha})q_{10}^2} \left(\eta - \left(\frac{q_{10}}{q_{00}}\right)^2\right) \geq 0, \\ \widetilde{\alpha} - \alpha_s &= = \frac{\overline{\alpha}(1-\underline{\alpha})q_{00}^2 q_{10}}{[\underline{\alpha} q_{00}^2 + (1-\underline{\alpha})q_{10}^2][\overline{\alpha} q_{00} + (1-\overline{\alpha})q_{10}]} \left(\eta - \frac{q_{10}}{q_{00}}\right) < 0. \end{aligned}$$

Therefore, $\overline{\alpha} \leq \widetilde{\alpha} < \alpha_s$. We consider the following two types of diagnosis-dependent patients.

Subcase 1(a): Initial illness perception satisfies $\underline{\alpha} < \alpha < \overline{\alpha}$. According to Proposition 1, the diagnosis-dependent patient in this range leaves whenever the two successive diagnosis results are consistent.

We start from a patient with illness perception $\overline{\alpha}_b < \alpha < \overline{\alpha}$. Under the worst-case scenario, she obtains a negative diagnosis at the first visit, and in the following visits, none of the diagnosis is consistent with the previous one. From (17), we conclude that she visits infinitely under the worst scenario. Based on it, we then have the following about the number of visits N:

1. If $N = 2i + 1$ $(i = 0, 1, 2, ...)$, it must be that the last diagnosis she receives is positive and all the rest $2i$ are "negative-positive" repeating i times $(i = 0, 1, 2, ...)$. Hence,

$$P(N = 2i + 1 | \overline{\alpha}_b < \alpha < \overline{\alpha}) = q_{11}(q_{11}q_{10})^i \alpha_b + q_{01}(q_{00}q_{01})^i (1 - \alpha_b);$$

2. If $N = 2i+2$ ($i = 0, 1, 2, ...$), it must be that the last two diagnoses both give negative results, and all diagnoses before the last two are negative-positive repeating i times ($i = 0, 1, 2, ...$); hence

$$P(N = 2i+2|\overline{\alpha}_b < \alpha < \overline{\alpha}) = (q_{10})^2(q_{11}q_{10})^i\alpha_b + (q_{00})^2(q_{00}q_{01})^i(1-\alpha_b).$$

That is, for $i = 0, 1, 2, ...$,

$$P(N|\overline{\alpha}_b < \alpha < \overline{\alpha}) = \begin{cases} q_{11}(q_{11}q_{10})^i\alpha_b + q_{01}(q_{00}q_{01})^i(1-\alpha_b) & \text{if } N = 2i+1, \\ (q_{10})^2(q_{11}q_{10})^i\alpha_b + (q_{00})^2(q_{00}q_{01})^i(1-\alpha_b) & \text{if } N = 2i+2. \end{cases} \tag{48}$$

Similarly, when $\underline{\alpha} < \alpha < \underline{\alpha}_b$, we can obtain that

$$P(N|\underline{\alpha} < \alpha < \underline{\alpha}_b) = \begin{cases} q_{10}(q_{11}q_{10})^i\alpha_b + q_{00}(q_{00}q_{01})^i(1-\alpha_b) & \text{if } N = 2i+1, \\ (q_{11})^2(q_{11}q_{10})^i\alpha_b + (q_{01})^2(q_{00}q_{01})^i(1-\alpha_b) & \text{if } N = 2i+2. \end{cases} \tag{49}$$

Subcase 1(b): Initial illness perception satisfies $\overline{\alpha} \leq \alpha < \alpha_s$. The patients whose initial illness perceptions fall into this range leave immediately if a positive result is observed at the first visit. When $\overline{\alpha} \leq \alpha < \widetilde{\alpha}$, if the patient obtains a negative result at the first visit,

$$\underline{\alpha}_b = g_0(\overline{\alpha}) < g_0(\alpha) < \overline{\alpha}_b;$$

thus, she pays at most two visits, and we show

$$P(N|\overline{\alpha} \leq \alpha < \widetilde{\alpha}) = \begin{cases} q_{11}\alpha_b + q_{01}(1-\alpha_b) & \text{if } N = 1, \\ q_{11}q_{10}\alpha_b + q_{00}q_{01}(1-\alpha_b) & \text{if } N = 2, +, \\ q_{10}^2\alpha_b + q_{00}^2(1-\alpha_b) & \text{if } N = 2, -, \end{cases} \tag{50}$$

where "+" and "−" denote that the second diagnosis is positive and negative, respectively. When $\widetilde{\alpha} \leq \alpha < \alpha_s$, if she observes a negative result at the first visit, since $g_0(\alpha) \geq g_0(\widetilde{\alpha}) = \overline{\alpha}_b$ and $g_0(\alpha) < g_0(\alpha_s) = \overline{\alpha}$, we show that her updated illness perception satisfies

$$\overline{\alpha}_b \leq g_0(\alpha) < \overline{\alpha};$$

thus, she behaves like the patients' whose illness perceptions are in interval $(\overline{\alpha}_b, \overline{\alpha})$ from the second visit on. We further show that for $i = 0, 1, 2, ...$,

$$P(N|\widetilde{\alpha} \leq \alpha < \alpha_s) = \begin{cases} q_{11}\alpha_b + q_{01}(1-\alpha_b) & \text{if } N = 1, \\ (q_{11}q_{10})^{i+1}\alpha_b + (q_{00}q_{01})^{i+1}(1-\alpha_b) & \text{if } N = 2i+2, \\ (q_{10})^3(q_{11}q_{10})^i\alpha_b + (q_{00})^3(q_{00}q_{01})^i(1-\alpha_b) & \text{if } N = 2i+3. \end{cases} \tag{51}$$

By using (46), we can further obtain the following:

$$\begin{aligned} E[N|\underline{\alpha} < \alpha < \underline{\alpha}_b] &= \frac{1+q_{11}}{1-q_{11}q_{10}}\alpha_b + \frac{1+q_{01}}{1-q_{00}q_{01}}(1-\alpha_b); \\ &= 1+\alpha_b\frac{q_{11}(1+q_{10})}{1-q_{11}q_{10}} + (1-\alpha_b)\frac{q_{01}(1+q_{00})}{1-q_{00}q_{01}} := 1+N_1, \\ E[N|\overline{\alpha}_b < \alpha < \overline{\alpha}] &= \frac{1+q_{10}}{1-q_{11}q_{10}}\alpha_b + \frac{1+q_{00}}{1-q_{00}q_{01}}(1-\alpha_b); \\ &= 1+\alpha_b\frac{q_{10}(1+q_{11})}{1-q_{11}q_{10}} + (1-\alpha_b)\frac{q_{00}(1+q_{01})}{1-q_{00}q_{01}} := 1+N_2, \\ E[N|\overline{\alpha} \le \alpha < \widetilde{\alpha}] &= 1+q_{10}\alpha_b + q_{00}(1-\alpha_b) := 1+N_3, \\ E[N|\widetilde{\alpha} \le \alpha < \alpha_s] &= 1+\alpha_b\frac{q_{10}(1+q_{10})}{1-q_{11}q_{10}} + (1-\alpha_b)\frac{q_{00}(1+q_{00})}{1-q_{00}q_{01}} := 1+N_4. \end{aligned} \tag{52}$$

Therefore,

$$\begin{aligned} E[N] =& 1-\Phi(\underline{\alpha}) + \left(\alpha_b\frac{q_{11}(1+q_{10})}{1-q_{11}q_{10}} + (1-\alpha_b)\frac{q_{01}(1+q_{00})}{1-q_{00}q_{01}}\right)[\Phi(\underline{\alpha}_b) - \Phi(\underline{\alpha})] \\ &+ \left(\alpha_b\frac{q_{10}(1+q_{11})}{1-q_{11}q_{10}} + (1-\alpha_b)\frac{q_{00}(1+q_{01})}{1-q_{00}q_{01}}\right)[\Phi(\overline{\alpha}) - \Phi(\overline{\alpha}_b)] \\ &+ [q_{10}\alpha_b + q_{00}(1-\alpha_b)][\Phi(\widetilde{\alpha}) - \Phi(\overline{\alpha}) \\ &+ \left(\alpha_b\frac{q_{10}(1+q_{10})}{1-q_{11}q_{10}} + (1-\alpha_b)\frac{q_{00}(1+q_{00})}{1-q_{00}q_{01}}\right)[\Phi(\alpha_s) - \Phi(\widetilde{\alpha})], \end{aligned}$$

Moreover, it can be shown that

$$N_2 - N_3 = q_{10}\alpha_b\left(\frac{1+q_{11}}{1-q_{11}q_{10}} - 1\right) + q_{00}(1-\alpha_b)\left(\frac{1+q_{01}}{1-q_{00}q_{01}} - 1\right) > 0; \tag{53}$$

$$N_4 - N_3 = q_{10}\alpha_b\left(\frac{1+q_{10}}{1-q_{11}q_{10}} - 1\right) + q_{00}(1-\alpha_b)\left(\frac{1+q_{00}}{1-q_{00}q_{01}} - 1\right) > 0. \tag{54}$$

Case 2: $\frac{q_{01}}{q_{11}} \le \eta < 1$. Under this case, when $\underline{\alpha} \le \alpha \le \overline{\alpha}_b$, the patient is obedient and visits only once; see Proposition 2. When $\overline{\alpha}_b < \alpha < \alpha_s$, the patient leaves if she obtains a positive result. If she obtains a negative result at the first visit, she becomes obedient because $\underline{\alpha} = g_0(\overline{\alpha}_b) < g_0(\alpha) < g_0(\alpha_s) = \overline{\alpha} \le \overline{\alpha}_b$, and hence she leaves regardless of the diagnosis result at the second visit. Therefore, the probability mass of her visiting times N is also given as (50), and her expected visiting times is given as (52). Therefore,

$$E[N] = 1 - \Phi(\underline{\alpha}) + [q_{10}\alpha_b + q_{00}(1-\alpha_b)][\Phi(\alpha_s) - \Phi(\overline{\alpha}_b)].$$

Proof of Proposition 4: From (43), (44), (45), and (47), we can show that

$$\frac{d\underline{\alpha}_b}{d\overline{\alpha}} = \frac{q_{11}q_{01}}{(\overline{\alpha}q_{01}+(1-\overline{\alpha})q_{11})^2} > 0, \tag{55}$$

$$\frac{d\overline{\alpha}_b}{d\underline{\alpha}} = \frac{q_{00}q_{10}}{(\underline{\alpha}q_{00}+(1-\underline{\alpha})q_{10})^2} > 0, \tag{56}$$

$$\frac{d\alpha_s}{d\overline{\alpha}} = \frac{q_{00}q_{10}}{(\overline{\alpha}q_{00}+(1-\overline{\alpha})q_{10})^2} > 0, \tag{57}$$

$$\frac{d\widetilde{\alpha}}{d\underline{\alpha}} = \left(\frac{q_{00}q_{10}}{\underline{\alpha}q_{00}^2+(1-\underline{\alpha})q_{10}^2}\right)^2 > 0. \tag{58}$$

When $\left(\frac{q_{01}}{q_{11}}\right)^2 \leq \eta < \frac{q_{01}}{q_{11}}$, by $\phi(\cdot) > 0$, (53), (54), (55), (56), (57) and (58), we can show that

$$\frac{\partial E[N]}{\partial \underline{\alpha}} = -(1+N_1)\phi(\underline{\alpha}) - N_2\phi(\overline{\alpha}_b)\frac{d\overline{\alpha}_b}{d\underline{\alpha}} - (N_4-N_3)\phi(\widetilde{\alpha})\frac{d\widetilde{\alpha}}{d\underline{\alpha}} < 0,$$

$$\frac{\partial E[N]}{\partial \overline{\alpha}} = N_1\phi(\underline{\alpha}_b)\frac{d\underline{\alpha}_b}{d\overline{\alpha}} + (N_2-N_3)\phi(\overline{\alpha}) + N_4\phi(\alpha_s)\frac{d\alpha_s}{d\overline{\alpha}} > 0.$$

When $\frac{q_{01}}{q_{11}} \leq \eta < 1$, by $\phi(\cdot) > 0$, (56), and (57), we can show that

$$\frac{\partial E[N]}{\partial \underline{\alpha}} = -\phi(\underline{\alpha}) - N_3\phi(\overline{\alpha}_b)\frac{d\overline{\alpha}_b}{d\underline{\alpha}} < 0, \quad \frac{\partial E[N]}{\partial \overline{\alpha}} = N_3\phi(\alpha_s)\frac{d\alpha_s}{d\overline{\alpha}} > 0. \tag{59}$$

Together with (41) and (42), we have

$$\frac{dE[N]}{dC_p} = \frac{\partial E[N]}{\partial \underline{\alpha}}\frac{d\underline{\alpha}}{dC_p} + \frac{\partial E[N]}{\partial \overline{\alpha}}\frac{d\overline{\alpha}}{dC_p} < 0. \tag{60}$$

For any given price f, the equilibrium arrival rate is determined by $\lambda = E[N]\Lambda$. By using $C_p = f + cw = f + \frac{c}{\mu-\lambda}$, we have $\frac{dC_p}{d\lambda} = \frac{c}{(\mu-\lambda)^2} = cw^2 > 0$. Then, together with (60), we have

$$\frac{dE[N]}{d\lambda} = \frac{dE[N]}{dC_p}\frac{dC_p}{d\lambda} < 0.$$

That is, $E[N]$ monotonically decreases in λ, and hence, the two functions $y = E[N]\Lambda$ (monotonically decreasing in λ) and $y = \lambda$ (monotonically increasing in λ) cross each other once. Therefore, there exists a unique equilibrium effective arrival rate.

Proof of Proposition 5: When doctor shopping is prohibited, $E[N] = 1 - \Phi(\underline{\alpha})$. Then, $\lambda = [1-\Phi(\underline{\alpha})]\Lambda$ and by (41),

$$\frac{d\lambda}{dC_p} = -\Lambda\phi(\underline{\alpha})\frac{d\underline{\alpha}}{dC_p} < 0.$$

From $C_p = f + cw$, we show

$$\frac{dC_p}{df} = 1 + c\frac{dw}{df} = 1 + \frac{c}{(\mu-\lambda)^2}\frac{d\lambda}{dC_p}\frac{dC_p}{df},$$

and hence, C_p increases with f because

$$0 < \frac{dC_p}{df} = \frac{1}{1 - cw^2 \frac{d\lambda}{dC_p}} < 1. \tag{61}$$

By (7), the subjective and objective rewards of a patient with illness perception $\alpha \notin S$ can be written as

$$\begin{aligned} E_{X_1}[r(X_1)\big|X_0 = \alpha] &= r(g_1(\alpha))P(s=1|\alpha) + r(g_0(\alpha))P(s=0|\alpha) = \kappa(\alpha), \\ E_{Y_1}[r(Y_1)\big|X_0 = \alpha] &= r(g_1(\alpha_b))P(s=1|\alpha_b) + r(g_0(\alpha_b))P(s=0|\alpha_b) = \kappa(\alpha_b), \end{aligned}$$

where $\kappa(\cdot)$ is given by (14). They are also the expected subjective and objective reward of an obedient patient. When $\frac{\widehat{\alpha}q_{00}}{\widehat{\alpha}q_{00}+(1-\widehat{\alpha})q_{10}} < \alpha < 1$, by (8), $g_0(\alpha) > \widehat{\alpha}$, and hence, the patient still prefers to seek treatment even upon receiving a negative diagnosis result. Therefore, when doctor shopping is not allowed,

$$\alpha_s^u = \frac{\widehat{\alpha}q_{00}}{\widehat{\alpha}q_{00} + (1-\widehat{\alpha})q_{10}}.$$

Recall that $\kappa(\cdot)$ is given by (14). We then obtain the average subjective reward and the average objective reward per patient as follows:

$$R_o^u = \int_{\underline{\alpha}}^{\alpha_s^u} \kappa(\alpha_b)\phi(\alpha)d\alpha, \quad R_s^u = \int_{\underline{\alpha}}^{\alpha_s^u} \kappa(\alpha)\phi(\alpha)d\alpha.$$

Recall the objective function given in (13):

$$\mathcal{W}_i^u = \Lambda\left(R_i^u - (1-\Phi(\underline{\alpha}))\frac{c}{\mu - (1-\Phi(\underline{\alpha}))\Lambda}\right),$$

where $i = o,\ s$. Then, for the objective social welfare, taking the first and second derivatives of the welfare function, we get

$$\begin{aligned} \frac{d\mathcal{W}_o^u}{d\underline{\alpha}} &= \Lambda\left(-\kappa(\alpha_b) + \frac{c\mu}{[\mu - [1-\Phi(\underline{\alpha})]\Lambda]^2}\right)\phi(\underline{\alpha}), \\ \frac{d^2\mathcal{W}_o^u}{d\underline{\alpha}^2} &= \Lambda\left[-\frac{2c\mu\Lambda}{[\mu - [1-\Phi(\underline{\alpha})]\Lambda]^3}f^2(\underline{\alpha}) + \left(-\kappa(\alpha_b) + \frac{c\mu}{[\mu - [1-\Phi(\underline{\alpha})]\Lambda]^2}\right)\phi'(\underline{\alpha})\right], \end{aligned}$$

where $\phi(\underline{\alpha}) > 0$. If there exists an $\underline{\alpha}^o$ such that

$$(1-\Phi(\underline{\alpha}^o))\Lambda = \mu - \sqrt{\frac{c\mu}{\kappa(\alpha_b)}},$$

then

$$\frac{d\mathcal{W}_o^u}{d\underline{\alpha}}\Big|_{\underline{\alpha}=\underline{\alpha}^o} = 0,\ \frac{d^2\mathcal{W}_o^u}{d\underline{\alpha}^2}\Big|_{\underline{\alpha}=\underline{\alpha}^o} < 0,$$

which indicates that $\mathcal{W}_o^u$ is unimodal with the global maximum at $\underline{\alpha} = \underline{\alpha}^o$. By (41) and (61), we can further show that

$$\frac{d\underline{\alpha}}{df} = \frac{d\underline{\alpha}}{dC_p}\frac{dC_p}{df} > 0. \tag{62}$$

That is, $\underline{\alpha}$ is strictly increasing in f, and hence, there is a one-to-one mapping between $\underline{\alpha}$ and f. Let $\tilde{f}$ be the price determined by $\underline{\alpha}^o$. Then, $\tilde{f}$ must be unique. Thus, if $\tilde{f} \geq 0$, $f_o^{u*} = \tilde{f}$ and the optimal arrival rate is $\lambda_o^u = \mu - \sqrt{\frac{c\mu}{\kappa(\alpha_b)}}$. Otherwise, $f_o^{u*} = 0$.

For the subjective social welfare, we consider the following two cases:

Case i. When $q_{11}V_1 - q_{10}L_1 \geq q_{00}V_0 - q_{01}L_0$, $\kappa'(\underline{\alpha}) \geq 0$.

$$\begin{aligned}
\frac{d\mathcal{W}_s^u}{d\underline{\alpha}} &= \Lambda\left(-\kappa(\underline{\alpha}) + \frac{c\mu}{[\mu - [1-\Phi(\underline{\alpha})]\Lambda]^2}\right)\phi(\underline{\alpha}),\\
\frac{d^2\mathcal{W}_s^u}{d\underline{\alpha}^2} &= \Lambda\Bigg[-\left(\kappa'(\underline{\alpha}) + \frac{2c\mu\Lambda\phi(\underline{\alpha})}{[\mu - [1-\Phi(\underline{\alpha})]\Lambda]^3}\right)\phi(\underline{\alpha})\\
&\qquad + \left(-\kappa(\underline{\alpha}) + \frac{c\mu}{[\mu - [1-\Phi(\underline{\alpha})]\Lambda]^2}\right)\phi'(\underline{\alpha})\Bigg],
\end{aligned}$$

where $\phi(\underline{\alpha}) > 0$. If there exists an $\underline{\alpha}^s$ such that

$$(1-\Phi(\underline{\alpha}^s))\Lambda = \mu - \sqrt{\frac{c\mu}{\kappa(\alpha_b)}},$$

then

$$\frac{d\mathcal{W}_s^u}{d\underline{\alpha}}\Big|_{\underline{\alpha}=\underline{\alpha}^s} = 0,\ \frac{d^2\mathcal{W}_s^u}{d\underline{\alpha}^2}\Big|_{\underline{\alpha}=\underline{\alpha}^s} < 0,$$

which indicates that $\mathcal{W}_s^u$ is unimodal with the global maximum at $\underline{\alpha} = \underline{\alpha}^s$. Recall from (62) that there is a one-to-one mapping between $\underline{\alpha}$ and f. Let $\hat{f}$ be the price determined by $\underline{\alpha}^s$. Then, $\hat{f}$ must be unique. Thus, iIf $\hat{f} \geq 0$, $f_s^{u*} = \hat{f}$, and hence, the optimal arrival rate is $\lambda_s^u = \mu - \sqrt{\frac{c\mu}{\kappa(\alpha_b)}}$. Otherwise, $f_s^{u*} = 0$.

Case ii. When $q_{00}V_0 - q_{01}L_0 > q_{11}V_1 - q_{10}L_1$, $\kappa'(\underline{\alpha}) < 0$. Since $\underline{\alpha} \leq \alpha_b$, we can show that

$$\frac{d\mathcal{W}_s^u}{d\underline{\alpha}} - \frac{d\mathcal{W}_o^u}{d\underline{\alpha}} = \Lambda\left(\kappa(\alpha_b) - \kappa(\underline{\alpha})\right) \leq 0.$$

Recall from the proof of Case i that $\mathcal{W}_o^u$ is unimodal, $\frac{d\mathcal{W}_o^u}{d\underline{\alpha}}\big|_{\underline{\alpha}=\underline{\alpha}^o} = 0$, and $\tilde{f}$ is the price determined by $\underline{\alpha}^o$. Also recall from 62 that $\underline{\alpha}$ increases in f. Thus, for those $f \geq \tilde{f}$, we have $\underline{\alpha} \geq \underline{\alpha}^o$ and $\frac{d\mathcal{W}_o^u}{d\underline{\alpha}} \leq \frac{d\mathcal{W}_o^u}{d\underline{\alpha}}\big|_{\underline{\alpha}=\underline{\alpha}^o} = 0$. Then, we have

$$\frac{d\mathcal{W}_s^u}{d\underline{\alpha}} \leq \frac{d\mathcal{W}_o^u}{d\underline{\alpha}} \leq 0,\ \forall f \geq \tilde{f}.$$

This indicates that $0 \leq f_s^{u*} \leq \max\{\tilde{f},\ 0\} = f_o^{u*}$.

Proof of Proposition 6: We have shown in Lemma 2 that η increases in C_p; together with (61), we show that

$$\frac{d\eta}{df} = \frac{d\eta}{dC_p}\frac{dC_p}{df} > 0. \tag{63}$$

And by (60) and (61), we have

$$\frac{d\lambda}{df} = \frac{d\lambda}{dC_p}\frac{dC_p}{df} = \Lambda\frac{dE[N]}{dC_p}\frac{dC_p}{df} < 0. \tag{64}$$

Recall from the proof of Proposition 5 that the obedient patients' objective rewards is $\kappa(\alpha_b)$. Below, we focus on the diagnosis-dependent patients.

Case 1: $\left(\frac{q_{10}}{q_{00}}\right)^2 < \eta < \frac{q_{10}}{q_{00}}$. When $\overline{\alpha}_b < \alpha < \overline{\alpha}$, given $i = 0, 1, ...$, if the number of visits paid by the patient $N = 2i+1$, we have $r(Y_N) = g_1 \cdot (g_1 \cdot g_0)^i(\alpha_b)V_1 - (1 - g_1 \cdot (g_1 \cdot g_0)^i(\alpha_b))L_0 = g_1(\alpha_b)V_1 - (1 - g_1(\alpha_b))L_0$; if $N = 2i + 2$ ($i = 0, 1, ...$), then we have $r(Y_N) = (1-(g_0)^2 \cdot (g_1 \cdot g_0)^i(\alpha_b))V_0 - (g_0)^2 \cdot (g_1 \cdot g_0)^i(\alpha_b)L_1 = (1-(g_0)^2(\alpha_b))V_0 - (g_0)^2(\alpha_b)L_1$. That is, when $\overline{\alpha}_b < \alpha < \overline{\alpha}$,

$$r(Y_N) = \begin{cases} g_1(\alpha_b)V_1 - (1 - g_1(\alpha_b))L_0 = \frac{q_{11}\alpha_b V_1 - q_{01}(1-\alpha_b)L_0}{q_{11}\alpha_b + q_{01}(1-\alpha_b)} & \text{if } N = 2i+1, \\ (1-(g_0)^2 \cdot (\alpha_b))V_0 - (g_0)^2 \cdot (\alpha_b)L_1 = \frac{q_{00}^2(1-\alpha_b)V_0 - q_{10}^2\alpha_b L_1}{q_{00}^2(1-\alpha_b)+q_{10}^2\alpha_b} & \text{if } N = 2i+2. \end{cases}$$

The corresponding probabilities of incurring N visiting time can be found in (48). We further show the following:

$$\begin{aligned} E_{Y_N}[r(Y_N)|\overline{\alpha}_b < \alpha < \overline{\alpha}] &= \sum_{i=0}^{\infty} \frac{q_{11}\alpha_b V_1 - q_{01}(1-\alpha_b)L_0}{q_{11}\alpha_b + q_{01}(1-\alpha_b)} P(N = 2i+1|\overline{\alpha}_b < \alpha < \overline{\alpha}) \\ &+ \sum_{i=0}^{\infty} \frac{q_{00}^2(1-\alpha_b)V_0 - q_{10}^2\alpha_b L_1}{q_{00}^2(1-\alpha_b) + q_{10}^2\alpha_b} P(N = 2i+2|\overline{\alpha}_b < \alpha < \overline{\alpha}) \\ &= \frac{q_{11}\alpha_b V_1 - q_{01}(1-\alpha_b)L_0}{1 - q_{11}q_{10}} + \frac{q_{00}^2(1-\alpha_b)V_0 - q_{10}^2\alpha_b L_1}{1 - q_{00}q_{01}}. \end{aligned}$$

Similarly, we show that when $\underline{\alpha} < \alpha < \underline{\alpha}_b$,

$$r(Y_N) = \begin{cases} (1 - g_0(\alpha_b))V_0 - g_0(\alpha_b)L_1 = \frac{q_{00}(1-\alpha_b)V_0 - q_{10}\alpha_b L_1}{q_{00}(1-\alpha_b)+q_{10}\alpha_b} & \text{if } N = 2i+1, \\ (g_1)^2 \cdot (\alpha_b)V_1 - (1-(g_1)^2 \cdot (\alpha_b))L_0 = \frac{q_{11}^2\alpha_b V_1 - q_{01}^2(1-\alpha_b)L_0}{q_{11}^2\alpha_b + q_{01}^2(1-\alpha_b)} & \text{if } N = 2i+2. \end{cases}$$

By using the corresponding probabilities of incurring N visiting time (49), we have

$$\begin{aligned} E_{Y_N}[r(Y_N)|\underline{\alpha} < \alpha < \underline{\alpha}_b] &= \sum_{i=0}^{\infty} \frac{q_{00}(1-\alpha_b)V_0 - q_{10}\alpha_b L_1}{q_{00}(1-\alpha_b) + q_{10}\alpha_b} P(N = 2i+1|\underline{\alpha} < \alpha < \underline{\alpha}_b) \\ &+ \sum_{i=0}^{\infty} \frac{q_{11}^2\alpha_b V_1 - q_{01}^2(1-\alpha_b)L_0}{q_{11}^2\alpha_b + q_{01}^2(1-\alpha_b)} P(N = 2i+2|\underline{\alpha} < \alpha < \underline{\alpha}_b) \\ &= \frac{q_{00}(1-\alpha_b)V_0 - q_{10}\alpha_b L_1}{1 - q_{00}q_{01}} + \frac{q_{11}^2\alpha_b V_1 - q_{01}^2(1-\alpha_b)L_0}{1 - q_{11}q_{10}}. \end{aligned}$$

When $\widetilde{\alpha} \le \alpha < \alpha_s$,

$$r(Y_N) = \begin{cases} g_1(\alpha_b)V_1 - (1-g_1(\alpha_b))L_0 = \frac{q_{11}\alpha_b V_1 - q_{01}(1-\alpha_b)L_0}{q_{11}\alpha_b + q_{01}(1-\alpha_b)} & \text{if } N = 1, \\ \alpha_b V_1 - (1-\alpha_b)L_0 & \text{if } N = 2i+2, \\ (1-(g_0)^3(\alpha_b))V_0 - (g_0)^3(\alpha_b)L_1 = \frac{q_{00}^3(1-\alpha_b)V_0 - q_{10}^3\alpha_b L_1}{q_{00}^3(1-\alpha_b)+q_{10}^3\alpha_b} & \text{if } N = 2i+3. \end{cases}$$

The corresponding probabilities of incurring N visiting time can be found in (51). We further obtain

$$\begin{aligned} E_{Y_N}[r(Y_N)|\widetilde{\alpha} \le \alpha < \alpha_s] &= \frac{q_{11}\alpha_b V_1 - q_{01}(1-\alpha_b)L_0}{q_{11}\alpha_b + q_{01}(1-\alpha_b)} P(N=1|\widetilde{\alpha} \le \alpha < \alpha_s) \\ &\quad + \sum_{i=0}^{\infty}[\alpha_b V_1 - (1-\alpha_b)L_0] P(N = 2i+2|\widetilde{\alpha} \le \alpha < \alpha_s) \\ &\quad + \sum_{i=0}^{\infty} \frac{q_{00}^3(1-\alpha_b)V_0 - q_{10}^3\alpha_b L_1}{q_{00}^3(1-\alpha_b)+q_{10}^3\alpha_b} P(N = 2i+3|\widetilde{\alpha} \le \alpha < \alpha_s) \\ &= q_{11}\alpha_b V_1 - q_{01}(1-\alpha_b)L_0 + \frac{q_{11}q_{10}[\alpha_b V_1 - (1-\alpha_b)L_0]}{1-q_{11}q_{10}} \\ &\quad + \frac{q_{00}^3(1-\alpha_b)V_0 - q_{10}^3\alpha_b L_1}{1-q_{00}q_{01}}. \end{aligned}$$

When $\overline{\alpha} \le \alpha < \widetilde{\alpha}$,

$$r(Y_N) = \begin{cases} g_1(\alpha_b)V_1 - (1-g_1(\alpha_b))L_0 = \frac{q_{11}\alpha_b V_1 - q_{01}(1-\alpha_b)L_0}{q_{11}\alpha_b + q_{01}(1-\alpha_b)} & \text{if } N = 1, \\ \alpha_b V_1 - (1-\alpha_b)L_0 & \text{if } N = 2,+, \\ (1-(g_0)^2\cdot(\alpha_b))V_0 - (g_0)^2\cdot(\alpha_b)L_1 = \frac{q_{00}^2(1-\alpha_b)V_0 - q_{10}^2\alpha_b L_1}{q_{00}^2(1-\alpha_b)+q_{10}^2\alpha_b} & \text{if } N = 2,-. \end{cases} \quad (65)$$

where "+" and "−" denote that the second diagnosis is positive and negative, respectively. The corresponding probabilities are given in (50), Thus,

$$\begin{aligned} E_{Y_N}[r(Y_N)|\overline{\alpha} \le \alpha < \widetilde{\alpha}] &= q_{11}\alpha_b V_1 - q_{01}(1-\alpha_b)L_0 + q_{11}q_{10}[\alpha_b V_1 - (1-\alpha_b)L_0] \\ &\quad + q_{00}^2(1-\alpha_b)V_0 - q_{10}^2\alpha_b L_1. \end{aligned}$$

Let

$$\varphi(\alpha) := E_{Y_N}[r(Y_N)|\alpha] - \kappa(\alpha_b),\ (\underline{\alpha},\ \underline{\alpha_b}) \cup (\overline{\alpha}_b, \alpha_s).$$

Hence, $\varphi(\alpha)$ represents the expected *objective reward improvements* of a diagnosis-dependent patient with initial illness perception α when doctor shopping is allowed. By plugging $q_{00} = q_{11} = q$ into the above equation and together with the above-obtained results, we then have

$$\varphi(\alpha) = \begin{cases} \frac{q_{00}q_{01}}{1-q_{00}q_{01}}[q(1-\alpha_b)(V_0+L_0) - (1-q)\alpha_b(V_1+L_1)] := \varphi_1, & \text{if } \underline{\alpha} < \alpha < \underline{\alpha}_b, \\ \frac{q_{00}q_{01}}{1-q_{00}q_{01}}[q\alpha_b(V_1+L_1) - (1-q)(1-\alpha_b)(V_0+L_0)] := \varphi_2, & \text{if } \overline{\alpha}_b < \alpha < \overline{\alpha}, \\ q_{00}q_{01}[\alpha_b(V_1+L_1) - (1-\alpha_b)(V_0+L_0)] := \varphi_3, & \text{if } \overline{\alpha} \le \alpha < \widetilde{\alpha}, \\ \frac{q_{00}q_{01}}{1-q_{00}q_{01}}[\alpha_b(V_1+L_1) - (1-\alpha_b)(V_0+L_0)] := \varphi_4, & \text{if } \widetilde{\alpha} \le \alpha < \alpha_s. \end{cases} \quad (66)$$

Obviously, $\varphi(\alpha)$ is piece-wise constant.

Case 2: $\frac{q_{10}}{q_{00}} \leq \eta < 1$. When $\overline{\alpha} \leq \alpha < \alpha_s$, the patient's reward is the same as $\overline{\alpha} \leq \alpha < \widetilde{\alpha}$ of Case 1; that is, her reward is the same as that in (65), and the corresponding probabilities is (50). Therefore, $\varphi(\alpha) = \varphi_3$.

We now look at the welfare maximization problem. By (52) and (56), we have $0 < N_3 < 1$ and $\frac{d\overline{\alpha}_b}{d\underline{\alpha}} > 0$. Then,

$$\kappa(\alpha_b) - \frac{c\mu N_3}{(\mu-\lambda)^2} > \kappa(\alpha_b) - \frac{c\mu}{(\mu-\lambda)^2}\left(1 + N_3 \frac{\phi(\overline{\alpha}_b)}{\phi(\underline{\alpha})}\frac{d\overline{\alpha}_b}{d\underline{\alpha}}\right) \geq 0 \text{ for } \lambda \leq \overline{\lambda}. \tag{67}$$

Recall that $\overline{f}$ is defined as the price that makes $\eta = q_{01}/q_{11}$. By (63), we show that for any $f \geq \overline{f}$, $\eta \geq q_{01}/q_{11}$. Therefore, when $f \geq \overline{f}$, the objective reward R_o then can be expressed as follows:

$$R_o = E_{Y_N}[\varphi(\alpha)|\overline{\alpha}_b < \alpha < \alpha_s] + \int_{\underline{\alpha}}^{\alpha_s} \kappa(\alpha_b)\phi(\alpha)d\alpha = E_{Y_N}[\varphi_3|\overline{\alpha}_b < \alpha < \alpha_s] + \int_{\underline{\alpha}}^{\alpha_s} \kappa(\alpha_b)\phi(\alpha)d\alpha,$$

where φ_3 is given in (66). Then, we have

$$\frac{\partial R_o}{\partial \underline{\alpha}} = -\varphi_3\phi(\overline{\alpha}_b)\frac{d\overline{\alpha}_b}{d\underline{\alpha}} - \kappa(\alpha_b)\phi(\underline{\alpha}),\ \frac{\partial R_o}{\partial \overline{\alpha}} = [\varphi_3 + \kappa(\alpha_b)]\,\phi(\alpha_s)\frac{d\alpha_s}{d\overline{\alpha}}. \tag{68}$$

Recall that $\lambda = E[N]\Lambda$. We can rewrite the objective function (12) as follows:

$$\max_f \mathcal{W}_o = \Lambda\left(R_o - E[N]\frac{c}{\mu-\lambda}\right) = \Lambda R_o - \frac{c\lambda}{\mu-\lambda} = \Lambda R_o + c - \frac{c\mu}{\mu-\lambda}.$$

By (59) and (68), when $f \geq \overline{f}$, we have

$$\begin{aligned}
\frac{d\mathcal{W}_o}{df} &= \left(\Lambda\frac{\partial R_o}{\partial \underline{\alpha}}\frac{d\underline{\alpha}}{dC_p} + \Lambda\frac{\partial R_o}{\partial \overline{\alpha}}\frac{d\overline{\alpha}}{dC_p} - \frac{c\mu}{(\mu-\lambda)^2}\frac{d\lambda}{dC_p}\right)\frac{dC_p}{df} \\
&= \Lambda\left[\left(\frac{\partial R_o}{\partial \underline{\alpha}} - \frac{c\mu}{(\mu-\lambda)^2}\frac{\partial E[N]}{\partial \underline{\alpha}}\right)\frac{d\underline{\alpha}}{dC_p} + \left(\frac{\partial R_o}{\partial \overline{\alpha}} - \frac{c\mu}{(\mu-\lambda)^2}\frac{\partial E[N]}{\partial \overline{\alpha}}\right)\frac{d\overline{\alpha}}{dC_p}\right]\frac{dC_p}{df} \\
&= -\Lambda\left[\kappa(\alpha_b)\phi(\underline{\alpha}) + \varphi_3\phi(\overline{\alpha}_b)\frac{d\overline{\alpha}_b}{d\underline{\alpha}} - \frac{c\mu}{(\mu-\lambda)^2}\left(\phi(\underline{\alpha}) + N_3\phi(\overline{\alpha}_b)\frac{d\overline{\alpha}_b}{d\underline{\alpha}}\right)\right]\frac{d\underline{\alpha}}{dC_p}\frac{dC_p}{df} \\
&\quad +\Lambda\left[\kappa(\alpha_b) + \varphi_3 - \frac{c\mu N_3}{(\mu-\lambda)^2}\right]\phi(\alpha_s)\frac{d\alpha_s}{d\overline{\alpha}}\frac{d\overline{\alpha}}{dC_p}\frac{dC_p}{df}.
\end{aligned}$$

Define

$$\overline{\lambda} := \mu - \sqrt{\frac{c\mu}{\kappa(\alpha_b)}\left(1 + N_3\frac{\phi(\overline{\alpha}_b)}{\phi(\underline{\alpha})}\frac{d\overline{\alpha}_b}{d\underline{\alpha}}\right)} = \mu - \sqrt{\frac{c\mu}{\kappa(\alpha_b)}\left(1 + \frac{q_{00}q_{10}[q_{10}\alpha_b + q_{00}(1-\alpha_b)]}{(\underline{\alpha}q_{00} + (1-\underline{\alpha})q_{10})^2}\frac{\phi(\overline{\alpha}_b)}{\phi(\underline{\alpha})}\right)}.$$

Then, if $\lambda|_{f=\overline{f}} \leq \overline{\lambda}$, by (64) we have that $\lambda \leq \lambda|_{f=\overline{f}} \leq \overline{\lambda}$ for $f \geq \overline{f}$. It is easy to show that that $\varphi_3 \geq 0$ when $0 < \widehat{\alpha} \leq \alpha_b$. By (41), (42), (56), (57), (61) and (67), we can show that when $0 < \widehat{\alpha} \le \alpha_b$ and $\lambda|_{f=\overline{f}} \le \overline{\lambda}$,

$$\frac{d\mathcal{W}_o}{df} < \Lambda\left[-\left(\kappa(\alpha_b) - \frac{c\mu}{(\mu-\lambda)^2}\left(1 + N_3\frac{\phi(\overline{\alpha}_b)}{\phi(\underline{\alpha})}\frac{d\overline{\alpha}_b}{d\underline{\alpha}}\right)\right)\phi(\underline{\alpha})\frac{d\underline{\alpha}}{dC_p} + \left(\kappa(\alpha_b) - \frac{c\mu N_3}{(\mu-\lambda)^2}\right)\phi(\alpha_s)\frac{d\alpha_s}{d\overline{\alpha}}\frac{d\overline{\alpha}}{dC_p}\right]\frac{dC_p}{df} < 0.$$

That is, when $\lambda|_{f=\overline{f}} \le \overline{\lambda}$ and $0 < \widehat{\alpha} \le \alpha_b$, $\mathcal{W}_o$ decreases with f for any $f \ge \overline{f}$, which indicates that if $\overline{f} \le 0$, $f_o^* = 0$, and if $\overline{f} > 0$, the maximum of $\mathcal{W}_o$ is found on the interval $[0,\ \overline{f}]$.

Appendix B Numerical Protocol: When Two Error Rates are Different

As shown in the proof of Lemma 1, the convexity of $v(x_n)$ does not depend on the assumption $q_{00} = q_{11}$. Hence, the double-threshold strategy still holds as long as the requirement stated in (9) is satisfied. Utilizing this property, we design the following numerical algorithm to find the system equilibrium outcome. The algorithm requires us to first discretize the state space (patients' belief) into multiple small intervals. We then calculate the expected visiting times for patients in each interval, based on which we can calculate the expected visiting times for the whole patient group. After that, the algorithm iterates between finding the effective arrival rate given the two thresholds and solving the dynamic program to find the two thresholds given the updated effective arrival rate.

The algorithm starts with the value of thresholds obtained under the One-Step Look-Ahead (OSLA) rule, which can be found in the proof of Proposition 1 and do not require $q_{00} = q_{11}$. That is, when $q_{00} \ne q_{11}$, the thresholds under the OSLA rule are exactly the same as those provided in the second lines of (10) and (11). That is, the initial value of two thresholds are

$$\underline{\alpha} = \frac{q_{01}(L_0+V_0)+C_p}{q_{11}(L_1+V_1)+q_{01}(L_0+V_0)} \text{ and } \overline{\alpha} = \frac{q_{00}(L_0+V_0)-C_p}{q_{00}(L_0+V_0)+q_{10}(L_1+V_1)},$$

where $C_p = f + c/(\mu-\lambda)$. For any given price f, we provide the following iterative algorithm to obtain the system equilibrium outcome.

Step 0 (Initialization): Set $v_0(x_n) = r(x_n)$, $\lambda = 0$, and $i = 1$. Let

$$\underline{s}(i) = \frac{q_{01}(L_0+V_0)+C_p}{q_{11}(L_1+V_1)+q_{01}(L_0+V_0)} \text{ and } \overline{s}(i) = \frac{q_{00}(L_0+V_0)-C_p}{q_{00}(L_0+V_0)+q_{10}(L_1+V_1)}.$$

Step 1: Let $\lambda^l = 0$ and $\lambda^h = \mu$.

Step 2: Set $\lambda = \frac{\lambda^l + \lambda^h}{2}$. Discretize the state space (patients' belief), calculate the expected visiting times of patients in each discretized state, and then obtain the expected visiting times $E[N]$ for the whole group of patients.

Step 3: If $\lambda - E[N]\Lambda < 0$, set $\lambda^l = \lambda$; otherwise, $\lambda^h = \lambda$.

Step 4: If $|\lambda - E[N]\Lambda| < 10^{-4}$, go to Step 5; otherwise, go to Step 2.

Step 5: Let $C_p = f + c/(\mu - \lambda)$ and

$$E\left[v_i(X_{n+1})|x_n\right] = P(s=0|x_n)v_{i-1}\left(g_0(x_n)\right) + P(s=1|x_n)v_{i-1}\left(g_1(x_n)\right).$$

Solve $E\left[v_i(X_{n+1})|x_n\right] - r(x_n) - C_p = 0$ with respect to x_n, and obtain two critical values x_n^1 and x_n^2 with $0 < x_n^1 < x_n^2 < 1$. Set $\underline{s}(i+1) = x_n^1$, $\overline{s}(i+1) = x_n^2$, and $i = i+1$.

Step 6: If $|\underline{s}(i) - \underline{s}(i-1)| < 10^{-4}$ and $|\overline{s}(i) - \overline{s}(i-1)| < 10^{-4}$, $\underline{\alpha} = \underline{s}(i)$ and $\overline{\alpha} = \overline{s}(i)$; otherwise, go to Step 1.