

Article

Contracts to Coordinate Healthcare Providers in the Telemedicine Referral System

Xianyi Wang ¹, Xiaofang Wang ¹ and Hui He ^{2,*}

¹ School of Business, Renmin University of China, Beijing 100872, China; 2016102585@ruc.edu.cn (X.W.); wangxiaofang@rmbs.ruc.edu.cn (X.W.)

² Business School, Beijing Technology and Business University, Beijing 100048, China

* Correspondence: herhui2006@126.com

Abstract: With the help of telemedicine, healthcare providers can increase patients' access to high-quality services while reducing the medical expenditure, especially for patients in remote areas. Once advanced care is needed, local patients will first be referred to an online health service and then be referred to the offline hospital if the online healthcare fails. In practice, local community hospitals and the advanced tertiary hospitals generally lack financial incentives to exert costly, but non-reimbursable, effort to avoid poor patient outcomes. Therefore, we build a new model to analyze the interaction between these two service providers, promoting them to exert the right effort by designing payment contracts. Our results show that neither fee-for-service nor bundled payment contracts can achieve the social optimum. Tertiary hospitals always exert less effort than the socially-optimal effort while the community hospital may exert less or more effort depending on the online treatment cost. Then, we propose a performance-based bundled payment contract that can coordinate both hospitals' decisions to achieve socially optimal outcomes. Finally, we numerically show the impact of the referral service fee and the online treatment cost on the efficiency of these contracts.

Keywords: contract design; telemedicine services; healthcare coordination; pay-for-performance; efficiency; payment system; community health



Citation: Wang, X.; Wang, X.; He, H. Contracts to Coordinate Healthcare Providers in the Telemedicine Referral System. *Sustainability* **2021**, *13*, 10299. <https://doi.org/10.3390/su131810299>

Academic Editor: Guillermo Palacios-Navarro

Received: 22 July 2021

Accepted: 10 September 2021

Published: 15 September 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Many primary care organizations are paid a fixed payment to manage patients' health conditions (i.e., Accountable Care Organizations, family doctors). They deliver the routine medical service and refer the patient for outside advanced care. However, such a system is not effective because many outside referrals are not necessary [1]. With the help of telemedicine, the primary care organization can treat more patients in the local area. Mohan et al. [2] show that by using the telemedicine service, less than 5% of patients need a referral to the diabetes center for further treatment, and 95% of the health problems can be managed locally.

In this paper, we consider a referral system consisting of a community hospital and a tertiary hospital that provides both online and offline services. The community hospital manages patients' health conditions and delivers routine care. Once advanced care is needed, the primary physician can communicate with the online physician of the tertiary hospital first. With the help of the online physician, patients may be cured locally. If the medical problem is solved after the online treatment procedure, patients will be discharged. Otherwise, they will be sent to the offline tertiary hospital for further care. An important consideration in this referral system is the effort levels at both medical institutions. To reduce the need for advanced care and avoid bad outcomes, the primary physician in the community hospital can invest in preventive care such as patient education and monitoring (the provision of information on the patient's behavioral and medical risks), clinical preventive services (immunization and vaccination of children, adults, and the elderly), population-based screening programs for early detection of diseases, dietary and

nutritional interventions in addition to the basic treatment. When a patient is referred to an online physician, the tertiary hospital exerts effort to improve the online outcome. The effort includes proper medication management (using consistent reminders to ensure that patients take medicine on time) and discharge management (communication and interactions with patients). Both hospitals' effort levels benefit social welfare but incur effort costs for each institution. Therefore, it is important to coordinate both hospitals and align their incentives with the goal of increasing social welfare and ensuring better patient outcomes.

Payment contracts play an important role in coordinating such effort levels. A common one is the fee-for-service (FFS) contract, where providers have a fixed payment for each service, regardless of the patient's health outcome. They are not financially responsible for the treatment failure or the complications. Additionally, more profits may be obtained due to the readmission. Some studies have found that the physician has no incentive to provide efficient, good health services under the FFS contract [3–5]. Thus, other payment contracts have been tested, including bundled payment (BP), shared savings, and the penalty contract. Some of these contracts' usefulness has been proven to be useful in coordinating participants' decisions. However, whether such contracts are helpful in the hospital–hospital setting, where both hospitals can make direct medical decisions to improve a patient's health status, is still unknown. Additionally, it is unclear who should pay the referral service fee. In most places in China, the local medical insurance department pays the referral service fee. While in some places (e.g., Shenzhen City, Guangdong Province; Dingyuan County, Anhui Province), the community hospital pays all the patient's medical costs, including services provided by outside health caregivers. In addition, the Chinese government encourages the medical reform that ensures that the community hospital receives the total medical insurance amount of patients and pays the service fee for outside referrals [6]. Meanwhile, we notice that some similar policies have been implemented in the US (i.e., the Affordable Care Act). Accountable Care Organizations are fully responsible for all medical costs of beneficiaries, including the outside referral cost.

In this paper, we focus on the payment design problem between the community hospital and the tertiary hospital, which provides both online and offline referral services. The objective of this paper is to analyze the performance of different payment contracts between two hospitals. Additionally, we investigate the impact of the fraction of the referral service fee paid by the community hospital for social welfare. We propose our research questions as follows: First, which types of contracts can coordinate both hospitals' effort levels to the social optimum? Second, what is the impact of cost-sharing between the community hospital and the government on the system's performance (i.e., social welfare, total cost)?

To address these questions, we consider a stylized model comprising the community hospital and the tertiary hospital. The community hospital pays a fraction of the referral service fee. The community hospital exerts effort to reduce the need for the outside service and the tertiary hospital exerts effort to improve the online treatment outcome. We compare the hospitals' effort decisions under the payment contract with the socially optimal effort decisions, which are determined by a central planner to maximize social welfare. Moreover, we analyze the impact of different parameters on social welfare and total cost under the payment contract.

This paper makes the following contributions. We propose a new model to capture the interaction of a community hospital, which exerts effort to manage the patient's health, and a tertiary hospital, which exerts effort to treat each referral online first and offers further offline service. Our results show that those payment contracts (i.e., FFS contract, BP contract) cannot achieve the social optimum. The tertiary hospital always exerts less effort than that required for the social optimum. The community hospital may exert less effort than the socially optimal effort when it pays a low fraction of referral service fees. Making the community hospital pay some referral service fee is beneficial for the total welfare, but may not be good for the total cost when the tertiary hospital's cost is low. To coordinate

the effort levels of both hospitals, we introduce a performance-based bundled payment contract, where the tertiary hospital receives a bundled payment and receives a penalty for the failed online treatment.

In the next section, we briefly review the related literature. After that, we introduce the framework of the proposed model. Based on the model, we set up a general framework of payment contracts and then analyze the performance of different contracts by varying parameters. Finally, we make suggestions for future research and offer some concluding remarks.

2. Literature Review

We classify the relevant research into three groups: (i) the contracting problem with joint effort decisions in supply chain management; (ii) telemedicine in the operations management (OM) area; (iii) the impact of payment contracts in health services.

A portion of the literature has studied the contracting problem with joint effort decisions in supply chain management [7–11]. Corbett and DeCroix [7] investigate several shared-savings contracts when both the buyer and the supplier can exert effort to reduce the product consumption. They find that such contracts can always increase supply chain profits, but do not lead to the reduced consumption. Chao et al. [9] investigate the contract design where the supplier and the manufacturer exert effort to improve their component failure rates. Kim and Netessine [11] analyze how contracting strategies influence supply chain parties' incentives to collaborate under information asymmetry. Unlike these papers, we consider the contracting problem with joint effort decisions in the healthcare area. The community hospital's effort influences the outside referral volume.

There is a growing body of literature investigating the management of telemedicine in the OM area. Rajan et al. [12] consider a primary care physician who manages the patient's office and e-visit services. When offering telemedicine and in-person modes, the specialist will set a lower price and serve more patients with a greater service rate. The total welfare will increase. Saghafian et al. [13] develop a model about the workload management in the telemedical physician triage and deploy it in a partially observable Markov decision process to describe the optimal policy. They find that the telemedical physician should evaluate more patients as the congestion in the emergency room increases. Unlike the above papers, Çakıcı and Mills [14] analyze the use of teletriage in managing healthcare demands and determine in which cases teletriage is effective and efficient. Zhong et al. [15] use Bayesian learning to model the opinion of online services. The severity of patients will be updated after the communication between the primary care physician and the specialist. In the work of Wang et al. [16], telemedicine plays an essential role in treating common and chronic diseases. The primary healthcare institution refers patients to the telemedicine firm (TF) or the general hospital (GH). Although increasing the referral rate to the TF can increase the profit of the TF and reduce the waiting time in the GH, the total utility of patients in the GH decreases greatly. The above papers do not consider designing payment schemes between service providers, which is our focus.

There have been some papers studying payment contracts in healthcare services [17–23]. Jiang et al. [24] investigate performance-based contracts in a principal-agent framework. The service provider maximizes his payoff by allocating his service capacity. Their results show that neither the fixed lump-sum payment nor FFS payment can achieve the socially optimal outcome. A linear performance-based contract is not even second best in the presence of asymmetric information. To overcome these shortcomings, they propose a threshold-penalty approach. Zhang et al. [25] focus on the payment scheme analysis in the Hospital Readmissions Reduction Program. Andritsos and Tang [5] analyze the effectiveness of different payment schemes. Both the patient and the hospital can jointly exert effort to reduce the patient's readmission. In their results, pay for performance (P4P) can help to reduce readmission more effectively than BP. However, BP outperforms P4P in lowering the cost. In the presence of co-payment, the system can reach optimum under either P4P or BP. Adida et al. [26] compare the performances of BP and FFS. Their results

confirm that the provider selects the excessive treatment level and the payer experiences a high cost under FFS. However, the physician chooses a less intensive treatment level under BP than that under the system optimum, which leads to a lower system payoff than that under FFS. By combining FFS and BP, they propose a new hybrid payment that improves most performance measures. Guo et al. [27] use a queuing model to capture the patient's readmission in a public healthcare system. They find that BP dominates FFS with a higher social welfare and a lower revisit rate when the patient pool size is large. Under FFS, patients experience a shorter waiting time. Rajagopalan and Tong [4] study the impact of different payment schemes on the coordination of care between the general physician and the specialist when both service providers are partially responsible for outcome costs. Their results show that the BP scheme brings higher referral rates, lower service time, and higher specialist utilization compared with an unbundled system. However, both payment schemes cannot achieve the social optimum. In addition, they introduce a two-part tariff scheme, in which the specialist receives a fixed price and compensation for each unit of service time, to coordinate the referral system. The above papers focus on the contract design between the payer and service provider. The payer makes the payment decision to influence the patient's health status indirectly. While in our paper, we focus on the payment contract between two hospitals.

The most related paper is that of Adida and Bravo [3]. They study contracts between an organization (requester) and an external provider (provider). The requester manages the health of a patient population and negotiates with the provider for the contract. They find that FFS cannot induce the social optimum. Unlike Adida and Bravo's work [3], our model framework incorporates the readmission from the online to the offline tertiary hospital. In addition, the tertiary hospital's profit is influenced by the online treatment failure. Furthermore, we consider the impact of the fraction of referral service fee related to the community hospital, while, in their work, the requester pays the full referral service fee.

3. Model Setting

We consider a healthcare system consisting of a tertiary hospital and a community hospital serving a given population of patients with a given condition (see Figure 1). People pay a fixed registration fee and sign the family doctor contract with the community hospital. They can freely receive basic medical care for a certain period. The payer (i.e., government) will pay fixed money to the community hospital for the management of patients. In this model, we focus on the contract design between the community hospital and the tertiary hospital. We describe a decision process for a chronic, non-emergency condition (such as hypertension, diabetes). The community hospital refers patients to the online physician of the tertiary hospital for advanced care. If the online treatment fails, the community hospital refers patients to the offline tertiary hospital for further treatment. As in many referral systems, we consider a setting where the community hospital arranges the referral service and there is no patient selection. We assume that the tertiary hospital will not deny service to a patient once they are contracted with the community hospital. Moreover, the tertiary hospital will send the patient to their offline department if the patient can not be cured online.

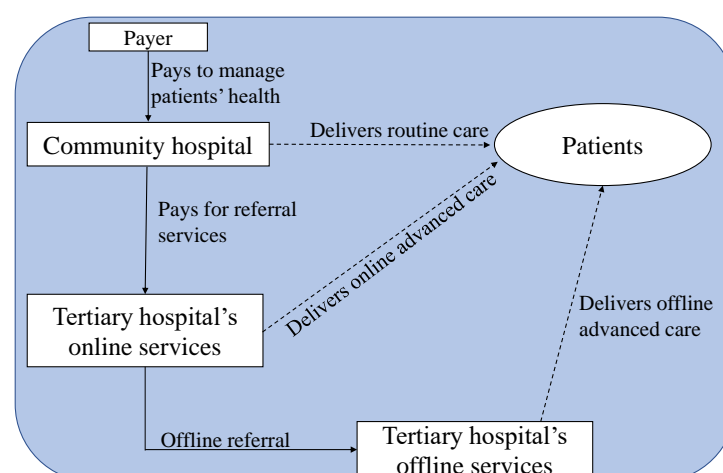


Figure 1. Interactions between participants in the telemedicine referral system.

3.1. Community Hospital

The community hospital delivers routine medical services to patients and manages their health. It bears the medical costs and pays a fraction of the outside referral service fee. To reduce the outside referral, the community hospital needs to invest in preventive measures such as patient education (the provision of information on behavioral and medical risks), clinical preventive services (immunization and vaccination of children, adults, and the elderly), population-based screening programs for early detection of diseases, dietary and nutritional intervention, and other methods facilitating care management. We summarize its investment as the effort level, e_c . Exerting the effort e_c incurs a cost, $C_c(e_c)$, for each patient. The size of the population served by the community hospital is exogenous and equal to λ_0 . The volume of outside referrals for advanced care is $\lambda(e_c)$, which is a function of the effort level selected by the community hospital. Adida et al. [3] provide a rationale for the negative impact of the increased effort spent with each patient on the outside referral, so we assume that $\lambda' < 0, \lambda'' > 0$ to reflect the diminishing effect of the community hospital's effort on the health outcome. The community hospital's decision on e_c reflects the following trade-off: Exerting more effort with the patient will result in better medical outcomes and fewer outside referrals, but it will lead to higher effort costs.

Although the community hospital can meet most of patients' needs, some outside referrals are inevitable (i.e., complex, specialized, or advanced procedures). If these happen, the patient will be referred to the online service provided by the tertiary hospital. The treatment cost is lower in the community hospital than in the tertiary hospital. Telemedicine services can reduce healthcare expenditure if the patient can be cured in the local area. Assisted by the online advice, patients do not need to travel a long distance to the offline tertiary hospital and experience the long waiting time for the service. So, it is beneficial for the community hospital to seek online medical advice first. The community hospital follows the online advice to treat the patient locally. Once this initial treatment fails, the patient will be referred to the offline tertiary hospital for further treatment.

3.2. Tertiary Hospital

The tertiary hospital provides both online and offline services. Once a patient is referred to the online physician, the tertiary hospital takes control of the patient's treatment program, including procedures beyond the standard one. The standard treatment cost of the online service is m_h . The tertiary hospital can make extra effort to improve the treatment outcome, such as using consistent reminders to ensure that patients take medicine on time, providing continuous communication and interactions with patients, monitoring patients' health status, and contacting the primary care physicians, etc. The tertiary hospital selects an effort level e_h and an effort cost $C_h(e_h)$ is incurred for each patient. Moreover, the initial online treatment may fail, and the patient needs to go to the offline tertiary hospital for

further treatment. The failed probability of online procedures is influenced by the tertiary hospital's effort. $q(e_h)$ denotes the probability that a patient can not be cured with the initial online treatment and requires further care in the offline tertiary hospital.

When the patient is referred to the offline tertiary hospital, an offline treatment cost M_h is incurred. After the offline treatment, the patient will be cured.

3.3. Payment Contract for the Referral Service

In reality, both hospitals' effort decisions may not be visible for the government and cannot be monitored all the time. For example, physicians can provide a dietary and nutritional intervention for the chronic-disease patient. The government does not know exactly the time spent and the intervention level for each patient. Additionally, the government cannot directly mandate that both hospitals always act in a socially optimal manner. Such a moral hazard problem has been frequently investigated in the literature [28–30]. For these reasons, the government will use the payment contract to incentivize the effort decisions of both hospitals indirectly. That is, the government will determine how each referral service is paid. Before analyzing the performance of payment contracts, we make some mild assumptions on the functions defined in the previous subsections.

Assumption 1. (Convexity of the effort cost, outside referral volume and the probability of online treatment failure).

- Effort costs $C_c(\cdot)$ and $C_h(\cdot)$ are non-negative and convex increasing functions.
- Outside referral volume $\lambda(\cdot)$ is a non-negative and convex decreasing function.
- The probability of online treatment failure $q(\cdot)$ is a non-negative and convex decreasing function.

The convex function of effort costs means that implementing basic actions (i.e., text, telephone follow-up treatment) incurs a low cost, while in order to improve the outcome, more costly actions should be taken. For example, physicians should record patients' daily blood pressure. They need to design software for patients to upload their blood pressure information and give the diet advice in time. Additionally, a free electric sphygmomanometer for people with high blood pressure can help them manage their health condition more effectively. The convexity of effort costs is a common assumption in the literature [31–33]. Furthermore, the volume of outside referrals is a convex decreasing function, which implies that a little increment in effort makes a big difference in the beginning. Giving regular follow-up calls to make sure patients take medicines correctly can reduce the probability of readmission. However, further effort become less influential. Similarly, we assume that the probability of treatment failure $q(\cdot)$ is a convex decreasing function in the tertiary hospital's effort. This assumption is consistent with common sense in the practice.

Assumption 2. Those functions satisfy the following conditions:

$$\frac{\lambda(e_c)\lambda''(e_c)}{(\lambda'(e_c))^2} \cdot \frac{q(e_h)q''(e_h)}{(q'(e_h))^2} \geq 1, \frac{\lambda(e_c)\lambda''(e_c)}{(\lambda'(e_c))^2} \cdot \frac{C_h(e_h)C_h''(e_h)}{(C_h'(e_h))^2} \geq 1.$$

4. Analysis

The two key decisions that impact patients' health conditions are effort levels selected by both hospitals. In this section, we first analyze the benchmark scenario where the central planner makes both effort decisions to achieve the social optimum. Then, we analyze the equilibrium under payment contracts where both hospitals make the effort decisions independently. Finally, we discuss an extension where the community hospital directly refers patients to the offline tertiary hospital.

4.1. Centralized System and the Social Optimum

We first consider the impact of outside referrals on patients. For each referral, patients have issues with things such as waiting time, which leads to discomfort. We define this discomfort in terms of a value loss. When patients are referred online, they experience a value loss of $-\mu_1$. If the online treatment fails and patients need to be referred offline, the additional value loss is $-\mu_2$, where $\mu_1, \mu_2 > 0$. Thus, the overall patient population's (dis)utility is

$$\Pi_{pu} = -\lambda(e_c)(\mu_1 + q(e_h)\mu_2).$$

A central planner will make both effort decisions to maximize social welfare of the healthcare system that considers patients, government and both hospitals. Additionally, we omit the revenue that the community hospital obtains from the family doctor service and the operation cost for routine care. The effort level does not impact these costs, so we treat them as constants.

The central planner optimizes the total social welfare for the entire system:

$$\max_{e_c, e_h \geq 0} \Pi(e_c, e_h) = -\lambda_0 C_c(e_c) - \lambda(e_c)(m_h + \mu_1 + q(e_h)(M_h + \mu_2) + C_h(e_h)),$$

where $\lambda(e_c)$ is the online referral volume. $\lambda_0 C_c(e_c)$ is the community hospital's total effort costs of the population. m_h is the treatment cost of the online referral service and $C_h(e_h)$ is the effort cost at the tertiary hospital. μ_1, μ_2 are the value loss of patients due to the referral service. $q(e_h)$ is the failure probability of the online treatment. M_h is the offline treatment cost.

We then have the following results on the optimal effort levels (e_c^*, e_h^*) at each hospital.

Proposition 1. *In the centralized system, the optimal effort levels (e_c^*, e_h^*) are obtained by solving the following equations:*

$$\begin{aligned} (M_h + \mu_2)q'(e_h) + C'_h(e_h) &= 0, \\ \lambda_0 C'_c(e_c) + \lambda'(e_c)(m_h + \mu_1 + q(e_h)(M_h + \mu_2) + C_h(e_h)) &= 0. \end{aligned}$$

Proof. See Appendix A. \square

We assume that $C'_h(0) = 0$ and $C'_c(0) = 0$ to guarantee optimal solutions are positive. It is consistent with the previous literature [7].

We now investigate the impact of system parameters on socially optimal outcomes.

Corollary 1. *Parameters M_h, m_h, μ_1, μ_2 have the following impact:*

- e_h^* increases in M_h, μ_2 and invariant in m_h, μ_1 ;
- e_c^* increases in m_h, M_h, μ_1, μ_2 .

Proof. See Appendix A. \square

These results are consistent with intuition. The social planner will exert more effort at the tertiary hospital when the offline treatment cost M_h and the value loss of the offline treatment μ_2 increase. As the tertiary hospital's effort can only affect the chance of the online treatment failure and the corresponding future cost, the social planner will not increase the effort level of the tertiary hospital. The tertiary hospital has no incentive to work harder when m_h, μ_1 increase. However, the central planner will exert more effort at the community hospital to avoid outside referrals when these parameters increase.

4.2. Decentralized System under Payment Contracts

Under a general payment contract framework, the community hospital first pays $\alpha\phi_1$ to the tertiary hospital for the telemedicine service, where ϕ_1 is the online service fee and $\alpha \in [0, 1]$ is the fraction of the referral service fee paid by the community hospital. The

remaining fraction of the service fee $(1 - \alpha)\phi_1$ will be paid by the government. However, with some probability, the initial treatment cannot cure the patient and an offline referral is needed. Once the offline referral happens, the community hospital will communicate with the tertiary hospital to make an appointment with the corresponding offline physician. Meanwhile, the community hospital should pay another service fee $\alpha\phi_2$ for the offline referral service. Similarly, ϕ_2 is the offline service fee. Both service fees ϕ_1, ϕ_2 are fixed and set in advance. In this section, the tertiary hospital's income includes both online and offline payments. We do not impose the assumption that the offline service income must cover the offline treatment cost. The offline service fee may be higher or lower than the offline treatment cost, and this assumption brings us more interesting insights.

Note that, in this general contract framework, the value of offline referral service fee ϕ_2 may be positive, zero, or negative. When the offline referral service fee is positive, the general payment contract is reduced to a traditional FFS payment contract. Under the FFS payment contract (that is, $\phi_1 > 0, \phi_2 > 0$), the tertiary hospital will receive the service fee for each referral service (either online or offline). When the offline referral service fee is zero (that is, $\phi_1 > 0, \phi_2 = 0$), the general payment contract is reduced to a BP contract where the tertiary hospital only receives a fixed payment ϕ_1 in advance for the entire treatment procedure, no matter whether the online treatment fails or not. The general payment contract is reduced to a pay-for-performance payment contract when the offline referral service fee is negative (that is, $\phi_1 > 0, \phi_2 < 0$). In this case, the tertiary hospital will receive a service fee ϕ_1 for the online treatment. However, if the online treatment fails, the tertiary hospital will have to pay a penalty ϕ_2 for the future treatment.

In this paper, we assume that the community hospital's objective is to jointly maximize its profit and its own patients' utility. This assumption is reasonable because the community hospital, as a basic medical organization, is public. Moreover, the family doctor policy is designed to inspire the community hospital to serve patients better and reward performance improvement. So, the community hospital should also consider its own profit. As the community hospital receives a fixed payment to manage patients' health and this payment is independent of the community hospital's effort decision, the community hospital's objective will focus on the cost. Additionally, we omit the revenue that the community hospital obtained from the family doctor service and the operation cost for routine care. The effort level does not impact these costs, so we treat them as constants and do not consider them in our model. Thus, given the tertiary hospital's effort level e_h , the community hospital decides the effort level e_c to optimize its decision

$$\max_{e_c \geq 0} \Pi_c(e_c; e_h) = -\lambda_0 C_c(e_c) - \lambda(e_c)(\alpha\phi_1 + \mu_1 + q(e_h)(\alpha\phi_2 + \mu_2)),$$

where $\alpha\phi_1 + q(e_h)\alpha\phi_2$ is the expected referral service fee paid by the community hospital. $\mu_1 + q(e_h)\mu_2$ is the expected patient's value loss of the referral service.

As Π_c is the combination of the cost of the community hospital and its own patients' utility, this objective function is not consistent with social welfare goal, which should include the objective of the tertiary hospital and the government expenditure.

The tertiary hospital decides the effort level e_h to maximize the profit:

$$\max_{e_h \geq 0} \Pi_h(e_h; e_c) = \lambda(e_c)(\phi_1 - m_h - C_h(e_h) + q(e_h)(\phi_2 - M_h)).$$

For each referral, $\phi_1 - m_h$ is the online treatment profit and $q(e_h)(\phi_2 - M_h)$ is the expected offline treatment profit.

To guarantee the tertiary hospital's participation, the service fees ϕ_1 and ϕ_2 should be high enough such that the tertiary hospital's optimal profit is non-negative when it exerts the optimal effort level e_h^G :

$$\phi_1 - m_h + q(e_h^G)(\phi_2 - M_h) \geq C_h(e_h^G). \quad (1)$$

We then have the following results on the optimal effort levels of both hospitals under the payment contract.

Proposition 2. *Under a payment contract,*

- if $\phi_2 \geq M_h$, the tertiary hospital's optimal effort $e_h^G = 0$, the community hospital's optimal effort e_c^G is the unique solution of

$$\lambda_0 C'_c(e_c) + \lambda'(e_c)(\alpha\phi_1 + \mu_1 + q(0)(\alpha\phi_2 + \mu_2)) = 0.$$

- if $\phi_2 < M_h$, the tertiary hospital's optimal effort e_h^G and the community hospital's optimal effort e_c^G are obtained by solving the following equations:

$$\begin{aligned} q'(e_h)(M_h - \phi_2) + C'_h(e_h) &= 0 \\ \lambda_0 C'_c(e_c) + \lambda'(e_c)(\alpha\phi_1 + \mu_1 + q(e_h)(\alpha\phi_2 + \mu_2)) &= 0. \end{aligned}$$

Proof. See Appendix A. \square

The following corollary shows how optimal effort levels of both hospitals change with the system parameters under the payment contract.

Corollary 2. *Parameters $M_h, m_h, \phi_1, \phi_2, \mu_1, \mu_2, \alpha$ have the following impact:*

- When $\phi_2 \geq M_h$, e_h^G is invariant in M_h, ϕ_2 ; otherwise, it increases in M_h and decreases in ϕ_2 ;
- Both e_c^G, e_h^G are invariant in m_h ;
- e_c^G always increases in $\alpha, \mu_1, \mu_2, \phi_1, \phi_2$. When $\phi_2 \geq M_h$, e_c^G is invariant in M_h ; otherwise, it decreases in M_h .

Proof. See Appendix A. \square

When the profit of the offline treatment is non-negative (i.e., $\phi_2 \geq M_h$), the tertiary hospital exerts no effort. In this situation, increasing the effort level at the tertiary hospital will reduce the online treatment failure probability and the future treatment cost partially paid by the community hospital. At the same time, the tertiary hospital will experience more effort costs and receives lower offline treatment profits. So, increasing the effort level at the tertiary hospital is costly to the tertiary hospital itself but benefits the community hospital. Additionally, the tertiary hospital can obtain extra profits if the patient is referred offline. So, the tertiary hospital does not change the effort level when M_h or ϕ_2 increases. However, when the offline treatment is non-profitable (i.e., $\phi_2 < M_h$), the tertiary hospital exerts more effort as the offline treatment cost M_h increases. When the offline referral service fee ϕ_2 increases, the tertiary hospital bears less profit loss for the offline treatment. So, it will decrease the effort level. As the tertiary hospital's effort only influences the probability of the online treatment failure and the future offline treatment cost, the tertiary hospital does not change the effort level when the online treatment cost m_h changes. In addition, the community hospital's effort is independent of the online treatment cost m_h .

The community hospital will exert more effort to reduce the outside referral when the referral service becomes more expensive (i.e., α, ϕ_1, ϕ_2 increase). Additionally, as patients experience more value loss for the outside referral service, the community hospital has more incentive to increase the effort level to avoid the referral. When the offline treatment is non-profitable, the tertiary hospital exerts more effort as the offline treatment cost M_h increases. The increased effort level at the tertiary hospital improves the outcome of the online treatment and reduces the future treatment cost. In this situation, the community hospital decreases its effort level.

In summary, when the offline treatment is profitable for the tertiary hospital, it exerts no effort. Otherwise, the tertiary hospital exerts more effort when the offline treatment cost is high, or the offline referral service fee is low. For the community hospital, it exerts more effort when the outside referral service becomes more costly. When the tertiary hospital

increases the effort level because of the increased offline treatment cost, the community hospital decreases its effort level.

4.3. Comparison of Effort Levels under Different Systems and Coordination to the Social Optimum

In this subsection, we compare the effort level (e_c^G, e_h^G) under a payment contract with the socially optimal outcomes. Consistent with the previous literature [27], we use the socially optimal outcomes as a benchmark. Thus, we define coordination as aligning both hospitals' effort levels with the social optimum.

Proposition 3. For all values of ϕ_1, ϕ_2 such that (1) holds,

- the tertiary hospital's effort $e_h^G \leq e_h^*$ if and only if $\phi_2 \geq -\mu_2$;
- if

$$m_h < q(e_h^G)\mu_2 - q(e_h^*)(M_h + \mu_2) - C_h(e_h^*),$$

$e_c^G > e_c^*$; otherwise, there exists a unique $\tilde{\alpha} \in [0, 1]$, such that $e_c^G \geq e_c^*$ when $\alpha \geq \tilde{\alpha}$, and $e_c^G < e_c^*$ when $\alpha < \tilde{\alpha}$. $\tilde{\alpha}$ is the unique solution of the equation

$$\lambda_0 C'_c(e_c^*) + \lambda'(e_c^*) (\alpha \phi_1 + \mu_1 + q(e_h^G)(\alpha \phi_2 + \mu_2)) = 0.$$

Proof. See Appendix A. \square

From Proposition 3, we know that the tertiary hospital exerts less effort than the socially optimal effort when the value of the offline referral service fee is high (i.e., $\phi_2 \geq -\mu_2$). We can explain it in two ways. When the profit of the offline treatment is non-negative (i.e., $\phi_2 \geq M_h$), exerting effort will only incur the additional cost and reduce the expected profit. The tertiary hospital has no incentive to exert any effort to improve the outcome. When treating patients offline is non-profitable (i.e., $\phi_2 < M_h$), the tertiary hospital exerts effort to reduce the need for the offline treatment. Different from the centralized system, the tertiary hospital considers the offline service profit $q(e_h)(\phi_2 - M_h)$, instead of $q(e_h)(M_h + \mu_2)$. From Corollary 2, we know that e_h^G decreases in ϕ_2 when $\phi_2 < M_h$. If the offline service fee ϕ_2 is high, the tertiary hospital experiences less profit loss and exerts less effort than the socially optimal effort. As the offline service fee ϕ_2 decreases, the tertiary hospital exerts more effort, and it finally exceeds the socially optimal effort when $\phi_2 < -\mu_2$.

From Corollary 2, we find that e_c^G increases in α . When the online treatment cost m_h is low, the expected referral cost per patient $(\alpha \phi_1 + \mu_1 + q(e_h^G)(\alpha \phi_2 + \mu_2))$ at the community hospital under the payment contract is always greater than under the centralized system, even if the community hospital pays no referral service fee (i.e., $\alpha = 0$). Therefore, the community hospital exerts more effort than the socially optimal effort to reduce more outside referrals. However, when the online treatment cost is high, the community hospital may exert less effort than the socially optimal effort. When the value of α is low (i.e., $\alpha < \tilde{\alpha}$), the community hospital experiences lower referral costs per patient under the payment contract than under the centralized system. So, it exerts less effort than the socially optimal outcome. As the community hospital bears a higher fraction of referral service fee, it increases the effort level to reduce the outside referral volume. When α is high (i.e., $\alpha > \tilde{\alpha}$), the referral cost per patient under the payment contract exceeds the referral cost under the centralized system. Thus, to curb these costs, the community hospital exerts more effort than the socially optimal effort.

It follows from Proposition 3 that there are no fees ϕ_1, ϕ_2 that may jointly coordinate both hospitals' effort levels under the FFS contract (i.e., $\phi_2 > 0$) or the BP contract (i.e., $\phi_2 = 0$). The tertiary hospital always exerts less effort than the socially optimal effort, while the community hospital may exert less effort or more effort than the socially optimal effort, depending on the value of the online treatment cost m_h .

From Proposition 3, we can select a penalty $\phi_2 = -\mu_2$ such that the tertiary hospital's effort level under the payment contract aligns with the socially optimal effort. Moreover,

the community hospital's effort may equal the socially optimal effort. Therefore, we define a new contract "performance-based bundled payment contract" which can ensure the coordination shown in the following proposition.

Proposition 4. A performance-based bundled payment contract with the online service fee $\phi_1^p = \frac{m_h + q(e_h^*)(M_h + \alpha\mu_2) + C_h(e_h^*)}{\alpha}$ and a penalty $\phi_2^p = -\mu_2$ can coordinate both hospitals' effort levels to achieve the socially optimal outcome.

Proof. See Appendix A. \square

The above result shows that the performance-based bundled payment contract can coordinate both hospitals' effort levels with the socially optimal effort. We make some observations. First, to align the tertiary hospital's effort with the socially optimal outcome, the community hospital should not pay any service fee for the offline treatment. Instead, it should charge a penalty for the online treatment failure. Second, the penalty is equal to the patient's value loss $-\mu_2$. By making the tertiary hospital take full responsibility for the treatment failure, the tertiary hospital will choose the socially optimal effort level. In this case, social welfare under the decentralized system is socially optimal.

4.4. The Impact of Cost-Sharing between the Community Hospital and the Government on the System's Performance

In this subsection, we analyze the impact of α on the system performance under the payment contract. We start with the analysis of both hospitals' profits and the total patients' utility. Based on these results, we further investigate the impact of α on social welfare and the total cost.

Proposition 5. $\Pi_c(e_c^G, e_h^G)$ decreases in α and $\Pi_h(e_c^G, e_h^G)$ is non-increasing in α . $\Pi_{pu}(e_c^G, e_h^G)$ increases in α .

Proof. See Appendix A. \square

As the community hospital pays a higher fraction of referral service fee, it exerts more effort to reduce the outside referral. However, the increase in the effort cost dominates the decrease in the referral cost, which leads to a lower profit. For the tertiary hospital, its profit decreases in α for the decreased referral volume when each referral profit margin is positive and is invariant in α if the profit margin is zero. However, the overall patient population's (dis)utility benefits from the increase in α due to the reduced outside referrals. When the government increases the fraction of referral service fees paid by the community hospital, both hospitals' profits will decrease. However, this leads to a lower value loss of patients.

In the following proposition, we analyze the impact of α on social welfare $\Pi(e_c^G, e_h^G)$.

Proposition 6. Under the payment contract, social welfare $\Pi(e_c^G, e_h^G)$ increases in α when $\phi_1 - m_h + q(e_h^G)(\phi_2 - M_h) = C_h(e_h^G)$; when $\phi_1 - m_h + q(e_h^G)(\phi_2 - M_h) > C_h(e_h^G)$, $\Pi(e_c^G, e_h^G)$ first increases and then decreases in α .

Proof. See Appendix A. \square

Under the payment contract, social welfare consists of the community hospital's optimal welfare $-\lambda_0 C_c(e_c^G) - \lambda(e_c^G)(\mu_1 + \alpha\phi_1 + q(e_h^G)(\alpha\phi_2 + \mu_2))$, the tertiary hospital's optimal profit $\lambda(e_c^G)(\phi_1 - m_h - C_h(e_h^G) + q(e_h^G)(\phi_2 - M_h))$ and the government's profit $-\lambda(e_c^G)(1 - \alpha)(\phi_1 + q(e_h^G)\phi_2)$. The community hospital's optimal welfare decreases in α and the tertiary hospital's optimal profit does not increase in α (Proposition 5). Additionally, the government's profit increases in α . When there is no profit for the tertiary hospital (i.e., $\phi_1 - m_h + q(e_h^G)(\phi_2 - M_h) = C_h(e_h^G)$), the impact of α on social welfare is determined by the

community hospital's optimal welfare and the government's profit. As the marginal welfare decrease of the community hospital, $-\lambda(e_c^G)(\phi_1 + q(e_h^G)\phi_2)$, is smaller than the marginal profit increase for the government, $\lambda(e_c^G)(\phi_1 + q(e_h^G)\phi_2) - \lambda'(e_c^G)(1 - \alpha)(\phi_1 + \alpha\phi_2)$ when α increases, social welfare increases in α . However, if the tertiary hospital gains a positive profit from the referral service, social welfare $\Pi(e_c^G, e_h^G)$ is not necessarily monotonic in α . It is determined by the marginal welfare decrease at the community hospital, the marginal profit decrease at the tertiary hospital and the marginal profit increase for the government. When the value of α is low, the government's profit increase dominates the welfare decrease at the community hospital and the profit decrease at the tertiary hospital as α increases. Hence, as the community hospital pays a higher fraction of the referral service fee, social welfare $\Pi(e_c^G, e_h^G)$ increases. However, for high α values, social welfare decreases in α : the decreased welfare and profit at both hospitals dominate any profit increase for the government.

From Proposition 6, we can observe that making the community hospital bear a referral service fee will improve social welfare. However, increasing the fraction of referral service fee at the community hospital may not always benefit the system. If the tertiary hospital gains no profit from the referral service, the community hospital paying the full referral service fee is better. When the tertiary hospital obtains positive profits, it is beneficial for the system to make the community hospital only pay a fraction of the referral service fee.

In the following, we analyze the impact of α on the total cost incurred by hospitals in the care delivery processes. We define the total cost as $\Pi_t(e_c, e_h) = \lambda_0 C_c(e_c) + \lambda(e_c)(m_h + q(e_h)M_h + C_h(e_h))$.

Proposition 7. *Under the payment contract, the total cost $\Pi_t(e_c^G, e_h^G)$ increases in α when $m_h + q(e_h^G)M_h + C_h(e_h^G) \leq \mu_1 + q(e_h^G)\mu_2$; otherwise, $\Pi_t(e_c^G, e_h^G)$ decreases first and then increases in α .*

Proof. See Appendix A. \square

From Corollary 2, we know that the community hospital's effort increases in α . As α does not impact the tertiary hospital's effort under the general payment contract, the cost at the tertiary hospital decreases for the reduced referral volume when α increases. As for the community hospital, the effort cost increases when α increases. When the online treatment cost at the tertiary hospital is low, the marginal cost increase at the community hospital dominates the cost decrease at the tertiary hospital, even if the community hospital pays no referral service fee (i.e., $\alpha = 0$). The total cost increases in α . However, if the online treatment cost at the tertiary hospital is high, the total cost decreases in α first and then increases. When the value of α is low, the cost decrease at the tertiary hospital dominates the cost increase at the community hospital as α increases. Hence, the total cost decreases when the community hospital pays a higher fraction of the referral service fee. However, for high α values, the increased cost at the community hospital dominates the decreased cost at the tertiary hospital. The total cost increases in α .

When the referral cost, $m_h + q(e_h^G)M_h + C_h(e_h^G)$, is low, the total cost reaches the minimum if the government pays the full fraction of referral service fee. Otherwise, by making the community hospital pay a fraction of referral service fee, the total cost reaches the minimum.

4.5. Extension: With or without Online Referrals

In this scenario, we assume there is no online treatment. The community hospital refers patients offline directly if the outside referral service is needed. As we assume that patients can be cured offline, the tertiary hospital makes no decision. For each referral, patients have issues with things such as waiting time which leads to a value loss $-\mu_2$ for the offline treatment.

When the outside referral happens, the community hospital pays a service fee $\alpha\phi_2$ to the tertiary hospital for this offline referral service. The community hospital decides the effort level e_c to optimize its decision:

$$\max_{e_c \geq 0} \Pi_c^N(e_c) = -\lambda_0 C_c(e_c) - \lambda(e_c)(\alpha\phi_2 + \mu_2),$$

In this case, we define the social welfare as $\Pi^N(e_c) = -\lambda_0 C_c(e_c) - \lambda(e_c)(M_h + \mu_2)$. Then, we obtain the following proposition.

Proposition 8. When $m_h \geq \frac{\lambda_0 C_c(e_c^N) + \lambda(e_c^N)(M_h + \mu_2) - \lambda_0 C_c(e_c^G)}{\lambda(e_c^G)} - \mu_1 - q(e_h^G)(M_h + \mu_2)$, the social welfare without online referrals, $\Pi^N(e_c^N)$, is higher than the social welfare with online referrals, $\Pi(e_c^G, e_h^G)$. e_c^N is the community hospital's optimal effort level in the offline referral system and is the unique solution of the following equation:

$$\lambda_0 C'_c(e_c) + \lambda'(e_c)(\alpha\phi_2 + \mu_2) = 0.$$

Proof. See Appendix A. \square

Proposition 6 shows that when the online treatment cost m_h is low, or the patient's value loss from online referral μ_1 or the failure rate of the online treatment $q(e_h^G)$ is low, introducing the online treatment of patients actually saves the costs (or equivalently, achieves a higher social welfare). This result is consistent with predictions and implies that hiring highly skilled online experts, increasing the efficiency of online treatment or reducing online treatment costs make it more socially beneficial to introduce online treatment before offline referrals.

5. Numerical Analysis

In Section 4, we proved that neither the FFS contract nor the BP contract can coordinate the effort decisions to the socially optimal outcome. However, these two contracts are commonly used in practice. Hence, we want to analyze the performances of these two contracts and obtain some managerial insights.

Let $C_c(e_c) = c_0^c e^{\beta e_c}$, $C_h(e_h) = c_0^h e^{\gamma e_h}$, $q(e_h) = q_0 e^{-\theta e_h}$ and $\lambda(e_c) = \lambda_0 e^{-\delta e_c}$. Parameters $\beta, \gamma, \theta, \delta$ are positive. These functions satisfy Assumptions 1 and 2. Then, we can obtain the nonborder solutions. The values of the parameters are set as $\beta = 0.1, \gamma = 0.2, \delta = 0.6, \theta = 0.2, c_0^c = 5, c_0^h = 15, q_0 = 0.3, \mu_1 = 20, \mu_2 = 100$. The online treatment cost is much more heterogeneous than the offline treatment cost as the service time of the initial treatment is more heterogeneous. Hence, we use two different values of $m_h \in \{20, 40\}$, to capture the wide possible range. Additionally, we set the offline treatment cost $M_h = 100$, the resident volume $\lambda_0 = 1000$ and the fraction of referral service fee $\alpha = 0.2$.

5.1. Fee-for-Service Payment Contract

In the FFS contract, we investigate a more typical case where $\phi_2 \geq M_h$ and reach the optimal solution (e_c^{FFS}, e_h^{FFS}) . We use the ratio $\rho^{FFS} = \frac{\Pi(e_c^{FFS}, e_h^{FFS})}{\Pi(e_c^*, e_h^*)}$ to show the inefficiency of the FFS contract in terms of social welfare, compared with the performance-based bundled payment contract. As the performance-based bundled payment contract achieves the maximal social welfare (equivalently, the minimum social cost, as Π is the negative social cost in our paper), this ratio is always greater than one. Additionally, the more inefficient the FFS is, the larger the value of the ratio. We set the value of ϕ_2 in the range $[M_h, 2M_h]$. In Figure 2 we show the impact of the referral service fee ϕ_2 and the online treatment cost m_h on the social welfare.

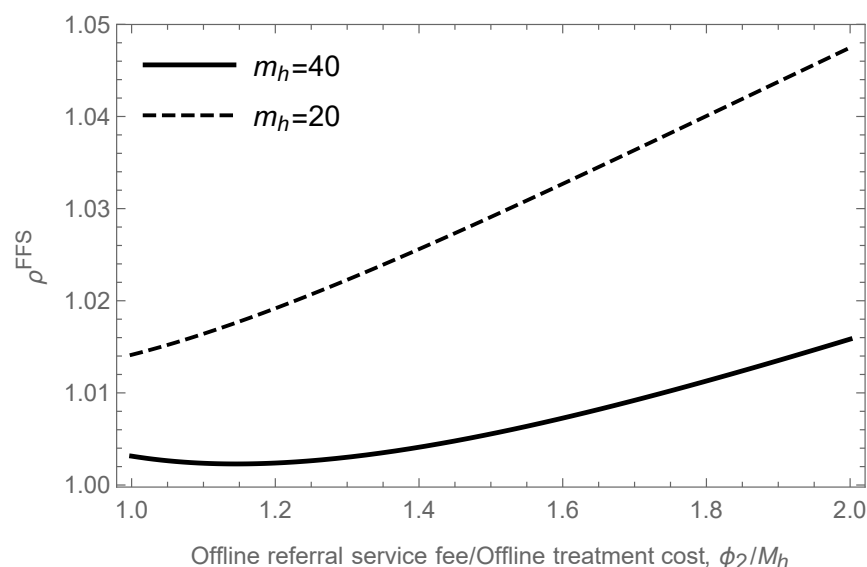


Figure 2. Inefficiency of the FFS contract.

Figure 2 shows that when the online treatment cost is low (e.g., $m_h = 20$), the FFS contract becomes more inefficient as the referral service fee ϕ_2 increases. This is because when serving an offline patient is profitable (i.e., $\phi_2 \geq M_h$), the tertiary hospital does not make any effort to improve the online treatment outcome. If the online treatment cost m_h is low, the community hospital exerts more effort than the socially optimal effort (Proposition 3). Because the community hospital's effort increases in ϕ_2 and the tertiary hospital's effort decision is independent of ϕ_2 , increasing ϕ_2 makes the community hospital's effort deviate further from the socially optimal effort and ultimately worsens the efficiency of the FFS contract. Interestingly, at a high value of the online treatment cost (e.g., $m_h = 40$), the inefficiency of the FFS contract first decreases and then increases in the referral service fee ϕ_2 . We can explain it as follows. When the online treatment cost m_h is high and the referral service fee ϕ_2 is low, the community hospital will exert less effort than the socially optimal effort. As ϕ_2 increases, the FFS contract becomes less inefficient and social welfare increases, which leads to a low ratio in the Figure 2. As the referral service fee ϕ_2 further increases, the community hospital's effort increases and exceeds the socially optimal effort. It makes social welfare deviate further from the social optimum.

It is important to note that the FFS contract is less inefficient when the online treatment cost m_h is higher (see Figure 2). Recall that from Corollary 1 we know that the socially optimal effort e_c^* increases in m_h . The community hospital's effort under the FFS contract is independent of m_h (Corollary 2). Hence, a high online treatment cost brings the socially optimal effort closer to that under the FFS contract, which increases the efficiency of the FFS contract.

We next analyze the impacts of the referral fee ϕ_2 and the online treatment cost m_h on the total cost $\Pi_t(e_c^{FFS}, e_h^{FFS})$. From Figure 3 we can see that the total cost increases for the offline referral service fee ϕ_2 . When the offline referral service becomes more expensive, the community hospital exerts more effort to reduce the outside referral volume. The increase in the effort cost dominates the decrease in the total referral cost, $\lambda(e_c^{FFS})(m_h + q(e_h^{FFS})M_h + C_h(e_h^{FFS}))$. So, the total cost increases. Additionally, when the online treatment cost m_h increases, the total cost increases. When the offline referral service fee ϕ_2 is low, decreasing the online treatment cost m_h will reduce the total cost more than when ϕ_2 is high.

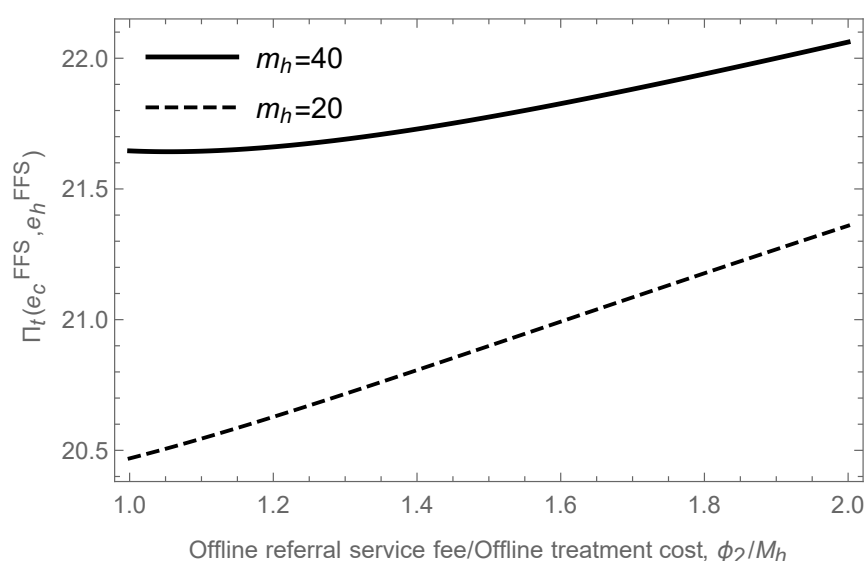


Figure 3. The impact of the referral fee ϕ_2 and the online treatment cost m_h on the total cost under the FFS contract.

5.2. Bundled Payment Contract

In the BP contract, we set the value of $\phi_2 = 0$ and reach the optimal effort levels (e_c^B, e_h^B) . We used the ratio $\rho^B = \frac{\Pi(e_c^B, e_h^B)}{\Pi(e_c^*, e_h^*)}$ to show the inefficiency of the BP contract in terms of social welfare, compared with the performance-based bundled payment contract. We varied the bundled payment ϕ_1 in the range $[m_h + C_h(e_h^B) + q(e_h^B)M_h, 2M_h]$.

Figure 4 illustrates that when the online treatment cost is low (e.g., $m_h = 20$), the BP contract becomes more inefficient as the bundled payment ϕ_1 increases. From Proposition 3, we know that the community hospital's effort is greater than the socially optimal effort when the online treatment cost is low. The tertiary hospital's effort is independent of the bundled payment ϕ_1 . The community hospital exerts more effort as the bundled payment ϕ_1 increases (Corollary 1). Hence, a higher bundled payment leads to a larger difference between the community hospital's effort under the BP contract and the socially optimal effort (see Figure 4). So, the ratio increases in ϕ_1 . Nevertheless, at a high value of the online treatment cost (e.g., $m_h = 40$), the inefficiency of the BP contract first decreases and then increases in the bundled payment ϕ_1 . This is because the community hospital exerts less effort than the socially optimal effort when ϕ_1 is low. As the community hospital's effort increases close to the socially optimal effort when ϕ_1 increases (see Corollary 2), the ratio decreases. If the bundled payment ϕ_1 is high, the community hospital's effort is higher than the socially optimal effort. Increasing ϕ_1 makes the community hospital's effort deviate further from the socially optimal effort and ultimately worsens social welfare. We can also see that a high value of online treatment cost m_h will make the bundled payment contract less inefficient.

Similarly, increasing the bundled payment ϕ_1 will increase the total cost (see Figure 5). When the value of the online treatment cost m_h is low, the increment in the total cost is faster than when the value of m_h is high, as the bundled payment ϕ_1 increases.

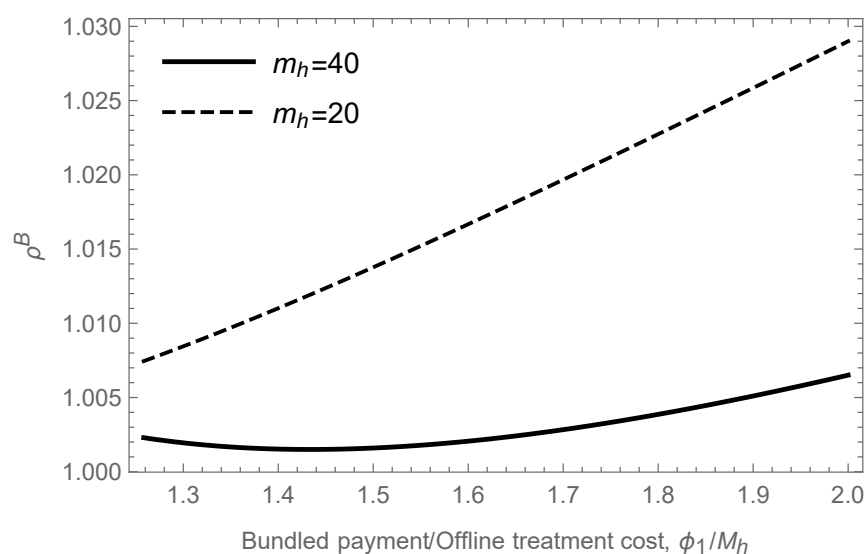


Figure 4. Inefficiency of the bundled payment contract.

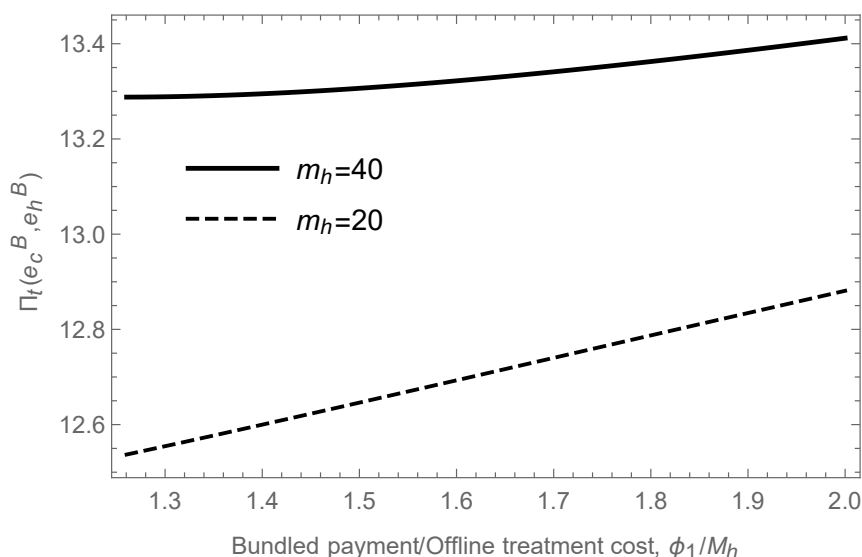


Figure 5. The impact of the bundled payment ϕ_1 and the online treatment cost m_h on the total cost under the BP contract.

6. Discussion

Cost, quality, and access, which are the “Iron Triangle” of healthcare, cannot be improved simultaneously in the traditional way. However, with the help of telemedicine, healthcare providers can increase the patient’s access to the high-quality service while reducing the medical expenditure, especially for patients in remote areas. Previous studies have focused on the impact of telemedicine on individual healthcare providers [12,34] or the referral system [15,16,35]. They have considered when to introduce telemedicine in the referral service system and whether the role of telemedicine is as a gatekeeper or the advanced care provider. However, they do not consider the payment scheme in the telemedicine referral system, which is an important decision by the healthcare providers and the policymaker. This paper analyzes the payment contract design problem between the community hospital and the tertiary hospital that provides both online and offline services.

Different from the previous research [24,26,27], we consider the payment contract design problem between two hospitals, instead of the payer and the caregiver. Additionally, the community hospital will bear a fraction of the referral service fee. Consistent with

previous research findings [3,4], our result shows that the FFS contract can not coordinate both hospitals' effort levels to be at the social optimum. Interestingly, social welfare may not always increase in terms of the fraction of referral service fee paid by the community hospital. This result is different from that of Adida et al. [3] and provides management insights for the policymaker to maximize social welfare. When the tertiary hospital gains no profit from the referral service, the government should have the community hospital pay the full referral service fee. Otherwise, there exists a unique optimal fraction. In addition, if the policymaker wants to minimize the total cost, when the referral cost per patient at the tertiary hospital is low, having the community hospital pay no referral service fee is always the best.

From what has been discussed above, we know that the efficiency of the FFS (BP) contract is bad when the referral service fee (bundled payment) is high and/or the online treatment cost is low. These results imply that for the treatment with the high referral service fee, the healthcare system's improvement is high from replacing the current payment contract with the performance-based bundled payment contract, especially when the online treatment cost is low.

7. Conclusions

The paper investigates the contract design problem between community hospitals and tertiary hospitals that provides both online and offline services. The community hospital receives a fixed payment from the payer to manage the health condition of a given population and, together with the government, pays the tertiary hospital for referral services. The payment contract between the community hospital and the tertiary hospital affects the incentives of both hospitals to exert effort to improve the treatment outcomes. Additionally, we consider the impact of the fraction of referral service fee paid by the community hospital on the system.

A variety of payment contracts between two hospitals have been analyzed in our paper. Our results confirm that the common payment contract (FFS, BP) lacks incentives to encourage the tertiary hospital to exert the socially optimal effort. The community hospital may exert less effort under the common payment contract (FFS, BP) than the socially optimal effort. Thus, neither contract can coordinate the referral system to the social optimum. However, adding a penalty to the tertiary hospital for the unsuccessful online treatment will make the decentralized system achieve the socially optimal outcome. Additionally, making the community hospital pay a fraction of referral service fee will improve social welfare, no matter under which contract.

There are some remaining questions. First, it will be interesting to examine contract design when the patient can choose the service. Second, we assume that the treatment costs are the same for all patients. In reality, they depend on patient severity. Patients with more severe diseases may incur higher costs for hospitals. Relaxing these assumptions will make the model richer but more complex, leading to interesting directions for future research.

Author Contributions: Conceptualization, X.W. (Xianyi Wang) and X.W. (Xiaofang Wang); methodology, X.W. (Xianyi Wang); software, X.W. (Xianyi Wang); validation, X.W. (Xiaofang Wang) and H.H.; formal analysis, X.W. (Xianyi Wang); writing—original draft preparation, X.W. (Xianyi Wang); writing—review and editing, X.W. (Xiaofang Wang) and H.H.; supervision, X.W. (Xiaofang Wang). All authors have read and agreed to the published version of the manuscript.

Funding: Hui He's work was supported by the Open Research Fund of Collaborative Innovation Centre for State-owned Assets Administration, Beijing Technology and Business University (grant number: BTBUGZGL201903).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

Proof of Proposition 1. The objective of the centralized system is

$$\Pi(e_c, e_h) = -\lambda_0 C_c(e_c) - \lambda(e_c)(m_h + \mu_1 + q(e_h)(M_h + \mu_2) + C_h(e_h)).$$

For the centralized system, we have that

$$\begin{aligned}\frac{\partial^2 \Pi}{\partial e_c^2} &= -\lambda_0 C_c'' - \lambda''(m_h + \mu_1 + q(M_h + \mu_2) + C_h). \\ \frac{\partial^2 \Pi}{\partial e_h^2} &= -\lambda((M_h + \mu_2)q'' + C_h''). \\ \frac{\partial^2 \Pi}{\partial e_c \partial e_h} &= -\lambda'((M_h + \mu_2)q' + C_h').\end{aligned}$$

Under Assumptions 1 and 2, $\frac{\partial \Pi^2}{\partial e_c^2} < 0$ and $\frac{\partial \Pi^2}{\partial e_h^2} < 0$. Thus, for the Hessian to be definite negative we need its determinant to be positive,

$$\underbrace{\lambda(\lambda_0 C_c'' + (m_h + \mu_1)\lambda'')((M_h + \mu_2)q'' + C_h'')}_{>0} + \underbrace{\lambda''\lambda((M_h + \mu_2)q + C_h)((M_h + \mu_2)q'' + C_h'') - (\lambda')^2((M_h + \mu_2)q' + C_h')^2}_{\Delta}.$$

Conditions in Assumption 2 guarantee $\Delta \geq 0$. Therefore, $\Pi(e_c, e_h)$ is jointly concave. So we can obtain the optimal decisions (e_c^*, e_h^*) from the following equations:

$$(M_h + \mu_2)q'(e_h) + C_h'(e_h) = 0, \quad (A1)$$

$$\lambda_0 C_c'(e_c) + \lambda'(e_c)(m_h + \mu_1 + q(e_h)(M_h + \mu_2) + C_h(e_h)) = 0. \quad (A2)$$

□

Proof of Corollary 1. From (A1) we can obtain that the optimal effort of the tertiary hospital, e_h^* , is only dependent on M_h, μ_2 and invariant in m_h, μ_1 . To see the variation with respect to M_h we can take total derivative and rearrange terms to get

$$q'(e_h^*) + ((M_h + \mu_2)q''(e_h^*) + C_h''(e_h^*)) \frac{\partial e_h^*}{\partial M_h} = 0.$$

Given the convexity of $C_h(\cdot)$ and $q(\cdot)$ together with $q(e_h)$ is decreasing in e_h , we have $\frac{\partial e_h^*}{\partial M_h} > 0$. Similarly, we take total derivative with respect to μ_2 . Then we have $\frac{\partial e_h^*}{\partial \mu_2} > 0$.

To analyze the community hospital effort, we use (A2) and take total derivative with respect to M_h . After rearranging terms, we get the following equality

$$\begin{aligned} & (\lambda_0 C_c''(e_c^*) + \lambda''(e_c^*)(m_h + \mu_1 + q(e_h^*)(M_h + \mu_2) + C_h(e_h^*))) \frac{\partial e_c^*}{\partial M_h} \\ & + \lambda'(e_c^*) \left((M_h + \mu_2)q'(e_h^*) \frac{\partial e_h^*}{\partial M_h} + C_h'(e_h^*) \frac{\partial e_h^*}{\partial M_h} + q(e_h^*) \right) = 0. \end{aligned}$$

From Assumption 1 and $(M_h + \mu_2)q'(e_h^*) + C_h'(e_h^*) = 0$ from the first-order condition, e_c^* is increasing in M_h . Similarly, we have $\frac{\partial e_c^*}{\partial \mu_2} > 0$.

We take total derivatives with respect to m_h ,

$$(\lambda_0 C_c''(e_c^*) + \lambda''(e_c^*)(m_h + \mu_1 + q(e_h^*)(M_h + \mu_2) + C_h(e_h^*))) \frac{\partial e_c^*}{\partial m_h} + \lambda'(e_c^*) = 0.$$

From Assumption 1, e_c^* is increasing in m_h . Similarly, we have $\frac{\partial e_c^*}{\partial \mu_1} > 0$. \square

Proof of Proposition 2. When $\phi_2 \geq M_h$, $\Pi_h(e_h; e_c)$ is decreasing in e_h under Assumptions 1 and 2. The optimal effort level is $e_h^G = 0$. Indeed, the tertiary hospital has no incentive to exert additional effort lower the chance of the treatment failure. That is costly and harms the community hospital's offline profit. Furthermore, under the Assumption 1, Π_c is concave in e_c . Therefore, the optimal solution e_c^G is obtained from the first-order condition:

$$\lambda_0 C_c'(e_c) + \lambda'(e_c)(\alpha\phi_1 + \mu_1 + q(0)(\alpha\phi_2 + \mu_2)) = 0. \quad (A3)$$

When $\phi_2 < M_h$, Π_h is concave in e_h and Π_c is concave in e_c under the Assumption 1. We can obtain the optimal solution (e_c^G, e_h^G) from the first-order conditions, according to the Assumptions 1 and 2:

$$q'(e_h)(M_h - \phi_2) + C_h'(e_h) = 0, \quad (A4)$$

$$\lambda_0 C_c'(e_c) + \lambda'(e_c)(\alpha\phi_1 + \mu_1 + q(e_h)(\alpha\phi_2 + \mu)) = 0. \quad (A5)$$

\square

Proof of Corollary 2. When $\phi_2 \geq M_h$, we take total derivative with respect to ϕ_1 in (A3),

$$\left(\lambda_0 C_c''(e_c^G) + \lambda''(e_c^G)(\alpha\phi_1 + \mu_1 + q(0)(\alpha\phi_2 + \mu_2)) \right) \frac{\partial e_c^G}{\partial \phi_1} + \alpha \lambda'(e_c^G) = 0.$$

From Assumption 1, we have that $\frac{\partial e_c^G}{\partial \phi_1} > 0$. Similarly, we have $\frac{\partial e_c^G}{\partial \phi_2} > 0$, $\frac{\partial e_c^G}{\partial \alpha} > 0$, $\frac{\partial e_c^G}{\partial \mu_1} > 0$, $\frac{\partial e_c^G}{\partial \mu_2} > 0$.

When $\phi_2 < M_h$, we take total derivative with respect to M_h in (A4),

$$\left(C_h''(e_h^G) + q''(e_h^G)(M_h - \phi_2) \right) \frac{\partial e_h^G}{\partial M_h} + q'_h(e_h^G) = 0.$$

From Assumption 1, we have $\frac{\partial e_h^G}{\partial M_h} > 0$. Similarly, we have $\frac{\partial e_h^G}{\partial \phi_2} < 0$.

To analyze the community hospital's effort, we use (A5) and take total derivative with respect to ϕ_2 . After rearranging terms, we get the following equality

$$\begin{aligned} & \left(\lambda_0 C_c''(e_c^G) + \lambda''(e_c^G)(\alpha\phi_1 + \mu_1 + q(e_h^G)(\alpha\phi_2 + \mu_2)) \right) \frac{\partial e_c^G}{\partial \phi_2} \\ & + \lambda'(e_c^G) \left(\alpha q(e_h^G) + q'(e_h^G)(\alpha\phi_2 + \mu_2) \frac{\partial e_h^G}{\partial \phi_2} \right) = 0. \end{aligned}$$

By the convexity Assumptions 1, 2 and $\frac{\partial e_h^G}{\partial \phi_2} < 0$, we have that $\frac{\partial e_c^G}{\partial \phi_2} > 0$. We also obtain

$$\begin{aligned} & \left(\lambda_0 C_c''(e_c^G) + \lambda''(e_c^G)(\alpha\phi_1 + \mu_1 + q(e_h^G)(\alpha\phi_2 + \mu_2)) \right) \frac{\partial e_c^G}{\partial M_h} \\ & + \lambda'(e_c^G) q'(e_h^G) \frac{\partial e_h^G}{\partial M_h} (\alpha\phi_2 + \mu_2) = 0. \end{aligned}$$

From Assumptions 1, 2 and $\frac{\partial e_h^G}{\partial M_h} > 0$, we have that $\frac{\partial e_c^G}{\partial M_h} < 0$. Because e_h^G is invariant in $\alpha, \phi_1, \mu_1, \mu_2$, we can obtain that $\frac{\partial e_c^G}{\partial \alpha} > 0, \frac{\partial e_c^G}{\partial \phi_1} > 0, \frac{\partial e_c^G}{\partial \mu_1} > 0, \frac{\partial e_c^G}{\partial \mu_2} > 0$, similar to the proof when $\phi_2 \geq M_h$. \square

Proof of Proposition 3. When $\phi_2 \geq M_h$, we have $e_h^G < e_h^*$.

When $\phi_2 < M_h$, from Corollary 2 we know that e_h^G is decreasing in ϕ_2 . So $e_c^G \leq e_c^*$ if and only if $M_h > \phi_2 \geq -\mu_2$.

We next compare the community hospital's effort. Let us denote the first order conditions for e_c under the general payment contract and the centralized system as

$$\begin{aligned}\psi^G(e_c, e_h) &= \lambda_0 C'_c(e_c) + \lambda'(e_c)(\alpha\phi_1 + \mu_1 + q(e_h)(\alpha\phi_2 + \mu_2)) \\ \psi(e_c, e_h) &= \lambda_0 C'_c(e_c) + \lambda'(e_c)(m_h + \mu_1 + C_h(e_h) + q(e_h)(M_h + \mu_2)).\end{aligned}$$

According to the Assumptions 1 and 2, $\frac{\partial \psi^G(e_c, e_h)}{\partial e_c} > 0$. We already know that $\psi^G(e_c^G, e_h^G) = 0$. If $\psi^G(e_c^*, e_h^G) \geq 0, e_c^G \leq e_c^*$, otherwise, $e_c^G > e_c^*$. From (A2) we know that $\psi(e_c^*, e_h^*) = 0$.

$$\begin{aligned}\psi^G(e_c^*, e_h^G) &= \lambda_0 C'_c(e_c^*) + \lambda'(e_c^*)(\alpha\phi_1 + \mu_1 + q(e_h^G)(\alpha\phi_2 + \mu_2)) \\ &= \psi^G(e_c^*, e_h^G) - \psi(e_c^*, e_h^*) \\ &= \lambda'(e_c^*)(\alpha\phi_1 - m_h - C_h(e_h^*) + q(e_h^G)(\alpha\phi_2 + \mu_2) - q(e_h^*)(M_h + \mu_2)).\end{aligned}$$

Because e_c^* and e_h^G are not influenced by α , we obtain that $\psi^G(e_c^*, e_h^G)$ is decreasing in α . When $\alpha = 1, \psi^G(e_c^*, e_h^G)$ reaches the minimum and e_c^G reaches the maximum from Corollary 2. In this situation, we have

$$\psi^G(e_c^*, e_h^G) = \lambda'(e_c^*)(\phi_1 - m_h - C_h(e_h^*) + q(e_h^G)(\phi_2 + \mu_2) - q(e_h^*)(M_h + \mu_2)).$$

Furthermore, by definition e_h^* minimizes the social cost over e_h and by convexity (Proposition 1) it is the unique minimizer of $C_h(e_h) + q(e_h)(M_h + \mu_2)$. Thus,

$$C_h(e_h^*) + q(e_h^*)(M_h + \mu_2) \leq C_h(e_h^G) + q(e_h^G)(M_h + \mu_2).$$

It follows that

$$\begin{aligned}\psi^G(e_c^*, e_h^G) &= \lambda'(e_c^*)(\phi_1 - m_h - C_h(e_h^*) + q(e_h^G)(\phi_2 + \mu_2) - q(e_h^*)(M_h + \mu_2)) \\ &\leq \lambda'(e_c^*)(\phi_1 - m_h - C_h(e_h^G) + q(e_h^G)(\phi_2 - M_h)) \leq 0,\end{aligned}$$

where the second inequality follows from (1). So $e_c^G \geq e_c^*$ when $\alpha = 1$.

When $\alpha = 0, \psi^G(e_c^*, e_h^G)$ reaches the maximum and e_c^G reaches the minimum. In this situation, we have

$$\psi^G(e_c^*, e_h^G) = \lambda'(e_c^*)(-m_h - C_h(e_h^*) + q(e_h^G)\mu_2 - q(e_h^*)(M_h + \mu_2)).$$

If $\psi^G(e_c^*, e_h^G) < 0$, we have $e_c^G > e_c^*$ for all $\alpha \in [0, 1]$; otherwise, there exists a unique $\tilde{\alpha}$ such that $e_c^G \leq e_c^*$ when $\alpha \leq \tilde{\alpha}$ and $e_c^G > e_c^*$ when $\alpha > \tilde{\alpha}$. $\tilde{\alpha}$ is the unique solution of the equation

$$\lambda_0 C'_c(e_c^*) + \lambda'(e_c^*)(\alpha\phi_1 + \mu_1 + q(e_h^G)(\alpha\phi_2 + \mu_2)) = 0.$$

\square

Proof of Proposition 4. From Proposition 3 we know that $e_h^G = e_h^*$ when $\phi_2^p = -\mu_2$. In this situation,

$$\lambda_0 C_c(e_c^G) + \lambda(e_c^G)(\alpha\phi_1 + \mu_1 + q(e_h^*)(-\alpha\mu_2 + \mu_2)) = 0.$$

Recall that

$$\lambda_0 C_c(e_c^*) + \lambda(e_c^*)(m_h + \mu_1 + q(e_h^*)(M_h + \mu_2) + C_h(e_h^*)) = 0.$$

According to Assumption 1, $\frac{d\Pi_c}{de_c}$ is decreasing in e_c . To ensure $e_c^* = e_c^G$, we have

$$\phi_1^p = \frac{m_h + q(e_h^*)(M_h + \alpha\mu_2) + C_h(e_h^*)}{\alpha}.$$

□

Proof of Proposition 5. We first analyze the impact of α on the community hospital's optimal welfare.

$$\begin{aligned} \frac{\partial \Pi_c(e_c^G, e_h^G)}{\partial \alpha} &= - \left(\lambda_0 C'_c(e_c^G) + \lambda'(e_c^G) (\alpha\phi_1 + \mu_1 + q(e_h^G)(\alpha\phi_2 + \mu_2)) \right) \frac{\partial e_c^G}{\partial \alpha} \\ &\quad - \lambda(e_c^G) (\phi_1 + q(e_h^G)\phi_2) \\ &= - \lambda(e_c^G) (\phi_1 + q(e_h^G)\phi_2) < 0, \end{aligned}$$

where the second equation come from

$$\lambda_0 C'_c(e_c^G) + \lambda'(e_c^G) (\alpha\phi_1 + \mu_1 + q(e_h^G)(\mu_2 + \alpha\phi_2)) = 0.$$

For the tertiary hospital's optimal profit

$$\Pi_h(e_c^G, e_h^G) = \lambda(e_c^G) (\phi_1 - m_h - C_h(e_h^G) + q(e_h^G)(\phi_2 - M_h))$$

is decreasing in α for the decreased referral volume $\lambda(e_c^G)$ when the marginal profit is positive. For patients' total utility $\Pi_{pu} = -\lambda(e_c^G)(\mu_1 + q(e_h^G)\mu_2)$, it is increasing in α for the decreased referral volume $\lambda(e_c^G)$. □

Proof of Proposition 6. The total social welfare evaluated at (e_c^G, e_h^G) is

$$\Pi(e_c^G, e_h^G) = -\lambda_0 C_c(e_c^G) - \lambda(e_c^G) (m_h + \mu_1 + q(e_h^G)(\mu_2 + M_h) + C_h(e_h^G)).$$

Because e_h^G is not influenced by α , we take the total derivatives with respect to α and obtain

$$\begin{aligned} \frac{\partial \Pi(e_c^G, e_h^G)}{\partial \alpha} &= - \left(\lambda_0 C'_c(e_c^G) + \lambda'(e_c^G) (m_h + \mu_1 + q(e_h^G)(\mu_2 + M_h) + C_h(e_h^G)) \right) \frac{\partial e_c^G}{\partial \alpha} \\ &= -\lambda'(e_c^G) (m_h - \alpha\phi_1 + q(e_h^G)(M_h - \alpha\phi_2) + C_h(e_h^G)) \frac{\partial e_c^G}{\partial \alpha}, \end{aligned}$$

where the second equation comes from

$$\lambda_0 C'_c(e_c^G) + \lambda'(e_c^G) (\alpha\phi_1 + \mu_1 + q(e_h^G)(\mu_2 + \alpha\phi_2)) = 0.$$

The sign of $\frac{\partial \Pi(e_c^G, e_h^G)}{\partial \alpha}$ is determined by $B(\alpha) = m_h - \alpha\phi_1 + q(e_h^G)(M_h - \alpha\phi_2) + C_h(e_h^G)$. $B(\alpha)$ is decreasing in α as $\phi_1 + q(e_h^G)\phi_2 > 0$ and $B(0) > 0$. $B(1) = m_h - \phi_1 + q(e_h^G)(M_h - \phi_2) + C_h(e_h^G) \leq 0$ from the constraint (1). $\frac{\partial e_c^G}{\partial \alpha} > 0$ from Corollary 2 and $\lambda'(e_h^G) < 0$. So if $\phi_1 - m_h + q(e_h^G)(\phi_2 - M_h) - C_h(e_h^G) = 0$, $\Pi(e_c^G, e_h^G)$ is increasing in $\alpha \in [0, 1]$; if $\phi_1 - m_h + q(e_h^G)(\phi_2 - M_h) - C_h(e_h^G) > 0$, $\Pi(e_c^G, e_h^G)$ is increasing in α first and then decreasing. □

Proof of Proposition 7. The total cost evaluated at (e_c^G, e_h^G) is

$$\Pi_t(e_c^G, e_h^G) = \lambda_0 C_c(e_c^G) + \lambda(e_c^G) \left(m_h + q(e_h^G) M_h + C_h(e_h^G) \right).$$

Because e_h^G is not influenced by α , we take the total derivatives with respect to α and obtain

$$\begin{aligned} \frac{\partial \Pi_t(e_c^G, e_h^G)}{\partial \alpha} &= \left(\lambda_0 C'_c(e_c^G) + \lambda'(e_c^G) \left(m_h + q(e_h^G) M_h + C_h(e_h^G) \right) \right) \frac{\partial e_c^G}{\partial \alpha} \\ &= \lambda'(e_c^G) \left(m_h - \alpha \phi_1 - \mu_1 + q(e_h^G) (M_h - \alpha \phi_2 - \mu_2) + C_h(e_h^G) \right) \frac{\partial e_c^G}{\partial \alpha}, \end{aligned}$$

where the second equation comes from

$$\lambda_0 C'_c(e_c^G) + \lambda'(e_c^G) \left(\alpha \phi_1 + \mu_1 + q(e_h^G) (\mu_2 + \alpha \phi_2) \right) = 0.$$

The sign of $\frac{\partial \Pi_t(e_c^G, e_h^G)}{\partial \alpha}$ is determined by $B(\alpha) - \mu_1 - q(e_h^G) \mu_2$. $B(\alpha)$ is decreasing in α and $B(1) - \mu_1 - q(e_h^G) \mu_2 < 0$ from the constraint (1). $\frac{\partial e_c^G}{\partial \alpha} > 0$ from Corollary 2 and $\lambda'(e_h^G) < 0$. So if $m_h - \mu_1 + q(e_h^G) (M_h - \mu_2) + C_h(e_h^G) \leq 0$, $\Pi_t(e_c^G, e_h^G)$ is increasing in $\alpha \in [0, 1]$; otherwise, $\Pi_t(e_c^G, e_h^G)$ is decreasing in α first and then increasing. \square

Proof of Proposition 8. Because the second-order condition is negative $\frac{d^2 \Pi_c^N(e_c)}{de_c^2} = -\lambda_0 C''(e_c) - \lambda''(e_c) (\alpha \phi_2 + \mu_2) < 0$, we can obtain the optimal solution from the first-order condition:

$$\frac{d \Pi_c^N(e_c)}{de_c} = -\lambda_0 C'_c(e_c) - \lambda'(e_c) (\alpha \phi_2 + \mu_2) = 0.$$

We can obtain the social welfare of the offline referral system $\Pi^N(e_c^N) = -\lambda_0 C_c(e_c^N) - \lambda(e_c^N) (\mu_2 + M_h)$. The difference of the social welfare between the offline referral system and the online referral system under the general payment contract is

$$\begin{aligned} \Pi(e_c^G, e_h^G) - \Pi^N(e_c^N) &= -\lambda_0 C_c(e_c^G) - \lambda(e_c^G) \left(m_h + \mu_1 + q(e_h^G) (M_h + \mu_2) \right) + \lambda_0 C_c(e_c^N) + \lambda(e_c^N) (M_h + \mu_2). \end{aligned}$$

Because e_c^G, e_h^G, e_c^N are invariant in m_h , we know that $\Pi(e_c^G, e_h^G) - \Pi^N(e_c^N)$ is decreasing in m_h . When $m_h \geq \frac{\lambda_0 C_c(e_c^N) + \lambda(e_c^N) (M_h + \mu_2) - \lambda_0 C_c(e_c^G)}{\lambda(e_c^G)} - \mu_1 - q(e_h^G) (M_h + \mu_2)$, $\Pi(e_c^G, e_h^G) \leq \Pi^N(e_c^N)$. \square

References

1. Lin, S.C.; Yan, P.L.; Moloci, N.M.; Lawton, E.J.; Ryan, A.M.; Adler-Milstein, J.; Hollingsworth, J.M. Out-Of-Network Primary Care Is Associated With Higher Per Beneficiary Spending In Medicare ACOs: An examination of the association between out-of-network care and per beneficiary spending using national Medicare data. *Health Aff.* **2020**, *39*, 310–318. [\[CrossRef\]](#)
2. Mohan, V.; Deepa, M.; Pradeepa, R.; Prathiba, V.; Datta, M.; Ravikumar, S.; Rakesh, H.; Sucharita, Y.; Webster, P.; Allender, S.; et al. Prevention of diabetes in rural India with a telemedicine intervention. *J. Diabetes Sci. Technol.* **2012**, *6*, 1355–1364. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Adida, E.; Bravo, F. Contracts for healthcare referral services: Coordination via outcome-based penalty contracts. *Manag. Sci.* **2019**, *65*, 1322–1341. [\[CrossRef\]](#)
4. Rajagopalan, S.; Tong, C. Payment Models to Coordinate Healthcare Providers with Partial Attribution of Outcome Costs. *Manuf. Serv. Oper. Manag.* **2021**. [\[CrossRef\]](#)
5. Andritsos, D.A.; Tang, C.S. Incentive programs for reducing readmissions when patient care is co-produced. *Prod. Oper. Manag.* **2018**, *27*, 999–1020. [\[CrossRef\]](#)

6. The State Council of the People's Republic of China. *Opinions of the General Office of the State Council on Reforming and Improving the Incentive Mechanism for Training and the Use of the General Practitioner*; The State Council of the People's Republic of China: Beijing, China, 2018. Available online: http://www.gov.cn/zhengce/content/2018-01/24/content_5260073.htm?trs=1 (accessed on 22 March 2021).
7. Corbett, C.J.; DeCroix, G.A. Shared-savings contracts for indirect materials in supply chains: Channel profits and environmental impacts. *Manag. Sci.* **2001**, *47*, 881–893. [\[CrossRef\]](#)
8. Corbett, C.J.; DeCroix, G.A.; Ha, A.Y. Optimal shared-savings contracts in supply chains: Linear contracts and double moral hazard. *Eur. J. Oper. Res.* **2005**, *163*, 653–667. [\[CrossRef\]](#)
9. Chao, G.H.; Iravani, S.M.; Savaskan, R.C. Quality improvement incentives and product recall cost sharing contracts. *Manag. Sci.* **2009**, *55*, 1122–1138. [\[CrossRef\]](#)
10. Jain, N.; Hasija, S.; Popescu, D.G. Optimal contracts for outsourcing of repair and restoration services. *Oper. Res.* **2013**, *61*, 1295–1311. [\[CrossRef\]](#)
11. Kim, S.H.; Netessine, S. Collaborative cost reduction and component procurement under information asymmetry. *Manag. Sci.* **2013**, *59*, 189–206. [\[CrossRef\]](#)
12. Rajan, B.; Tezcan, T.; Seidmann, A. Service systems with heterogeneous customers: Investigating the effect of telemedicine on chronic care. *Manag. Sci.* **2019**, *65*, 1236–1267. [\[CrossRef\]](#)
13. Saghaian, S.; Hopp, W.J.; Iravani, S.M.; Cheng, Y.; Diermeier, D. Workload management in telemedical physician triage and other knowledge-based service systems. *Manag. Sci.* **2018**, *64*, 5180–5197. [\[CrossRef\]](#)
14. Çakıcı, Ö.E.; Mills, A.F. On the Role of Telerriage in Healthcare Demand Management. *Manuf. Serv. Oper. Manag.* **2020**. [\[CrossRef\]](#)
15. Zhong, X.; Xu, A.; He, Q.; Prakash, A.M. Modeling and Analysis of E-Consults in Primary-Specialty Care Referrals. *IEEE Trans. Autom. Sci. Eng.* **2020**, *18*, 311–322. [\[CrossRef\]](#)
16. Wang, X.; Zhang, Z.; Yang, L.; Zhao, J. Price and capacity decisions in a telemedicine service system under government subsidy policy. *Int. J. Prod. Res.* **2020**, *59*, 5130–5143. [\[CrossRef\]](#)
17. Lee, H.H.; Pinker, E.J.; Shumsky, R.A. Outsourcing a two-level service process. *Manag. Sci.* **2012**, *58*, 1569–1584. [\[CrossRef\]](#)
18. Hennig-Schmidt, H.; Selten, R.; Wiesen, D. How payment systems affect physicians' provision behaviour—An experimental investigation. *J. Health Econ.* **2011**, *30*, 637–646. [\[CrossRef\]](#) [\[PubMed\]](#)
19. Allard, M.; Jelovac, I.; Leger, P.T. Treatment and referral decisions under different physician payment mechanisms. *J. Health Econ.* **2011**, *30*, 880–893. [\[CrossRef\]](#)
20. Selviaridis, K.; Wynstra, F. Performance-based contracting: A literature review and future research directions. *Int. J. Prod. Res.* **2015**, *53*, 3505–3540. [\[CrossRef\]](#)
21. Lu, M.; Albert Ma, C.T.; Yuan, L. Risk selection and matching in performance-based contracting. *Health Econ.* **2003**, *12*, 339–354. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Kantarevic, J.; Kralj, B. Physician payment contracts in the presence of moral hazard and adverse selection: The theory and its application in Ontario. *Health Econ.* **2016**, *25*, 1326–1340. [\[CrossRef\]](#)
23. Nguyen, H.T.H.; Bales, S.; Wagstaff, A.; Dao, H. Getting incentives right? The impact of hospital capitation payment in Vietnam. *Health Econ.* **2017**, *26*, 263–272. [\[CrossRef\]](#) [\[PubMed\]](#)
24. Jiang, H.; Pang, Z.; Savin, S. Performance-based contracts for outpatient medical services. *Manuf. Serv. Oper. Manag.* **2012**, *14*, 654–669. [\[CrossRef\]](#)
25. Zhang, D.J.; Gurvich, I.; Van Mieghem, J.A.; Park, E.; Young, R.S.; Williams, M.V. Hospital readmissions reduction program: An economic and operational analysis. *Manag. Sci.* **2016**, *62*, 3351–3371. [\[CrossRef\]](#)
26. Adida, E.; Mamani, H.; Nassiri, S. Bundled payment vs. fee-for-service: Impact of payment scheme on performance. *Manag. Sci.* **2017**, *63*, 1606–1624. [\[CrossRef\]](#)
27. Guo, P.; Tang, C.S.; Wang, Y.; Zhao, M. The impact of reimbursement policy on social welfare, revisit rate, and waiting time in a public healthcare system: Fee-for-service versus bundled payment. *Manuf. Serv. Oper. Manag.* **2019**, *21*, 154–170. [\[CrossRef\]](#)
28. Hoppe, E.I.; Schmitz, P.W. Contracting under incomplete information and social preferences: An experimental study. *Rev. Econ. Stud.* **2013**, *80*, 1516–1544. [\[CrossRef\]](#)
29. Aghion, P.; Fudenberg, D.; Holden, R.; Kunitomo, T.; Tercieux, O. Subgame-perfect implementation under information perturbations. *Q. J. Econ.* **2012**, *127*, 1843–1881. [\[CrossRef\]](#)
30. Hoppe, E.I.; Schmitz, P.W. Hidden action and outcome contractibility: An experimental test of moral hazard theory. *Games Econ. Behav.* **2018**, *109*, 544–564. [\[CrossRef\]](#)
31. Ren, Z.J.; Zhou, Y.P. Call center outsourcing: Coordinating staffing level and service quality. *Manag. Sci.* **2008**, *54*, 369–383. [\[CrossRef\]](#)
32. Li, C. Sourcing for supplier effort and competition: Design of the supply base and pricing mechanism. *Manag. Sci.* **2013**, *59*, 1389–1406. [\[CrossRef\]](#)
33. Roels, G. Optimal design of coproductive services: Interaction and work allocation. *Manuf. Serv. Oper. Manag.* **2014**, *16*, 578–594. [\[CrossRef\]](#)

-
34. Zhong, X.; Hoonakker, P.; Bain, P.A.; Musa, A.J.; Li, J. The impact of e-visits on patient access to primary care. *Health Care Manag. Sci.* **2018**, *21*, 475–491. [[CrossRef](#)]
 35. Wang, X.; Zhang, Z.; Zhao, J.; Shi, Y. Impact of telemedicine on healthcare service system considering patients' choice. *Discret. Dyn. Nat. Soc.* **2019**. [[CrossRef](#)]