Service System Design of Video Conferencing Visits with Nurse Assistance

Xiaojie Wang
Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, xwang828@ufl.edu

Yongpei Guan
Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, guan@ise.ufl.edu

Xiang Zhong
Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL 32611, xiang.zhong@ise.ufl.edu

Despite providing convenience and reducing the travel burden of patients, video-conferencing (VC) clinical visits haven't enjoyed the wide uptake by patients and care providers. It is desired that the medical problems addressed by VC visits can match a face-to-face encounter in scope and quality. Subsequently, VC visits with nurse assistance are emerging; however, the scalable and financially sustainable of such services are unclear. Therefore, we explore the implementability of VC visits with nursing services using a game-theoretic model, and investigate the impact of different pricing schemes (discriminative pricing based on patient characteristics vs. non-discriminative) on patients' care choices between VC and in-person visits. Our results shed light on the "artificial congestion" created by a profit-driven medical institution that hurts patient welfare, and subsequently identify the conditions where the interest of the social planner and the medical institution are aligned. Our results highlight that, compared to a uniform price of VC visits which seems fair, discriminative pricing can be more beneficial for patients and the medical institution alike. This heightens the importance of insurance coverage of telehealth related services to promote the adoption of telehealth by patients and care providers, and ultimately, improving care access and patient outcomes.

Key words: Telehealth; nurse assistance; game-theoretic; pricing scheme; artificial congestion

1. Introduction

Video-conferencing (VC) visits, as one type of telehealth services, have been offered to ambulatory patients to manage their care, and have garnered growing attention. Studies of VC visits have demonstrated its benefits in a spectrum of clinical settings, including wound care, prenatal genetic screening, cardiovascular care, and home care (Abrams and Geier 2006, Clegg et al. 2011, Grant et al. 2015, Eriksson et al. 2011). Systematic reviews found that VC visits were associated with decreased travel costs and lost time/wages, increased access to social support, and a better ability to tailor care delivery to patient and family needs (Sevean et al. 2009). An American Well study showed that 20 % of consumers would switch their current primary care provider to another who offers telehealth services in their area (American Well 2019a). The global market of telehealth was anticipated to expand at a compound annual growth rate of 14.3 % from 2014 to 2020, based on a report by Nathaniel Lacktman, Esq (Wood 2019).

Despite the above-mentioned advantages and a growing market, several factors limit the wide adoption of VC visits, such as infrastructure, technology literacy, and privacy, among which, the ability of clinicians to perform an adequate physical examination during VC visits was primarily concerned (Powell et al. 2017, CDC 2020). The companion of a qualified medical personnel during the VC visits is desired to empower telehealth to cover more disease conditions (Kitamura et al. 2010). The regular services rendered during an office visit, like vital sign check and basic physical examination, can be conducted by a nurse at patient homes with no compromise in care quality (Allen et al. 1995), and sending nurses for home care of patients with cancer has been demonstrated feasible (Bohnenkamp et al. 2004). Combining virtual and in-person care is the path forward of telehealth (AmwellHealth 2020). However, the cost of dispatching nurses to all patient homes is prohibitive. The payment policy regarding the VC visits with nurse assistance has not been established. Charging the same price for auxiliary nursing services to patients from different communities might be fair; however, it is unclear if this "fairness" comes at a price. There is a lack of best practice for medical institutions to follow. Hence, how to integrate VC visits with traditional office visits to meet patient needs while achieving scalable and financially sustainable operations warrants a thorough investigation.

To fill this gap, we develop a game-theoretic framework to investigate the optimal pricing strategies that navigate patients between VC visits and in-person visits, thereby achieving an overall economic outcome and improving patient care. The medical institution or social planner is modeled as the leader in the Stackelberg game, which determines whether to offer VC visits to supplement in-person visits, and the corresponding prices charged for auxiliary nursing services. Patients from multiple communities as followers make decisions to maximize their perceived utility. Models featuring different pricing schemes (discriminative pricing based on patient characteristics vs. non-discriminative) and efficient algorithms for solving the optimal pricing strategy under both linear and general piece-wise linear concave nurse cost functions are provided. These modeling efforts lead to the following major observations and policy insights:

(i) Our model sheds light on the value of telehealth in diverting patients from in-person visits to virtual visits that reduces excessive travel burden and congestion in the system, and thus improving the overall system efficiency. The key determinator is the "system-level" marginal gain from VC visits as a function of individual travel burden and nurse coordination costs. In principle, the medical institution favors patients from larger communities that are moderately far away from its central care facility, indicating a larger marginal gain, to receive VC visits. For patients, there exists a threshold-type equilibrium patient diversion strategy. Our model further reveals the characteristics of patients' care choices in response to the medical institution's pricing strategy as well as payment and cost structures. Surprisingly, the existence of the threshold-type equilibrium is invariant to these changes.

(ii) Hospital congestion is generated when too many patients conduct in-person visits, which adds negative externality (e.g., a higher infection risk during a pandemic) to the service system. However, it is not in the medical institution's best interests to completely eliminate congestion. The medical institution is less sensitive to congestion compared to patients, and the perception of congestion at the central care facility will induce patients to favor VC visits and thus being willing to pay a higher price for "customized" and "hassle-free" care. This "artificial congestion" created by the medical institution is not desired by the social planner, but our model reveals the existence of a pricing regime that aligns the interest of the social planner and the medical institution, which bolsters the successful implementation of VC visits.

(iii) Our results also highlight that, charging a uniform price of VC visits, which seems fair to patients, can lead to unintended consequences. The conditions that favor the implementation of VC visits become more restrictive, and patients from different communities are disproportionally affected. Patients from the communities that are not very far from the central hospital and are close to satellite clinics (low nurse coordination cost) would have access to VC visits under discriminative pricing. However, they would lose this option because the medical institution is not willing to provide it under non-discriminative pricing. In addition, the aggregate patient utility and hospital revenue can be lower when discretionary prices are forbidden. Therefore, a discriminative pricing strategy can be more beneficial. Besides lifting the fairness constraint, shifting the burden from patients to the payer (e.g., let the government or insurance companies share cost) can further facilitate the adoption of telehealth by patients and the medical institution alike.

The remainder of the paper is organized as follows: The related literature is briefly reviewed in Section 2. The assumptions of the game-theoretic model are described in Section 3. The analyses under the discriminative and non-discriminative pricing schemes are presented in Section 4 and Section 5, respectively. Extensions and a numerical study using real-world inspired data are introduced in Sections 6 and 7. Finally, concluding remarks are given in Section 8.

2. Literature Review

In the realm of telehealth research, the mainstream literature splits into clinical studies and health economics studies. Existing clinical studies on the use of VC visits mainly focused on the patient perception of and their experience with VC visits, as well as the evidence of effectiveness (Kitamura et al. 2010, Mallow et al. 2016). On cost-effectiveness, home telehealth services were found to lead to reductions in the costs of health care resources for chronic diseases such as congestive heart failure, diabetes, and chronic obstructive pulmonary disease, etc., from both health care systems' and insurers' perspectives (see review papers Polisena et al. (2009) and Dávalos et al. (2009)). These studies considered the cost associated with specific health care resources, such as hospitalizations, primary care encounters, and emergency department visits, and also included the cost required to set up a home telehealth system. A 34% monthly cost reduction after using virtual care services was revealed in one study (Grady 2002), where the primary attributes were the reduction in hospitalization (by 14%), along with reduced appointment times (5-10 minutes), and the removal of considerable travel time for both patients and their provider. The initial installation cost of a virtual care system was estimated at a rate of $720 per month, but the reduction of other costs was expected to cover the differences.

The operations engineering society has limited works on but shows a growing interest in telehealth. A variety of analytical models have been developed. For instance, mixed integer programming was used to find the best telemedicine device in a telemedicine workstation in rural communities in South Africa (Treurnicht 2009). Simulation techniques were used in the design of a telemedicine program in Mexico, which was formulated as a vehicle routing problem with a mobile unit equipped with telecommunication gear and satellite connection (Lach and Vázquez 2004). In addition, a Bayesian network was used to enhance the telehealth system design by predicting any problematic situation for at-home kidney disease patients (Bellot et al. 2002). Meta-heuristics was used to develop a genetic algorithm based method to allow for the tele-screening of breast cancer using digital mammography images (Qian et al. 2005).

Meanwhile, stochastic models and game theory were prevalently used in the design and evaluation of flexible service systems with strategic entities. Relevant literature includes an optimal tele-specialist policy designed to recommend which patients to treat remotely considering the quality and accuracy trade-off of telehealth services (Tarakci et al. 2007). Telehealth physician triage as a hierarchical knowledge-based service system has been analyzed using a partially observable Markov decision process to describe the optimal scheduling policy (Saghafian et al. 2018). The impact of electronic visits (e-visits) on the cycle time of office visits in primary care settings was investigated using a multi-class vacation queue model in Zhong et al. (2017, 2018). How physicians select the size of their patient panel and patient revisit intervals to maximize their compensation under the e-visit model using patient health dynamics and Markov decision process were presented in Bavafa et al. (2018) and Bayram et al. (2020). On telehealth service design, the specialist's optimal service rate and price for a telehealth service considering patients differing in their travel distances were investigated in Rajan et al. (2019). Models of telehealth as on-demand service platforms to investigate pricing and service rate decisions can be found in Liu et al. (2018) and Savin et al. (2019). Furthermore, the conditions under which switching from an office visit-based clinical practice to a mobile-based practice is economical was investigated in Rajan and Agnihothri (2019). Despite these efforts, the service system design of VC visits with nurse assistance and the corresponding operational challenges have not been addressed.

3. Problem Settings

Large health care systems typically maintain a network of multiple facilities, including medical centers, hospitals, and outpatient sites, serving a large population of patients with care of varying complexity. Our work is motivated by the telehealth service design for such systems to deliver outpatient care (e.g., specialty care). Without loss of generality, we use the University of Florida Health Cancer Center (UFHCC), a community hospital, to motivate the study. The UFHCC serves a catchment area of 22 counties in North Central Florida, with the farthest one being a three-hour drive away from the medical center in Gainesville, FL (see Figure 1). Among patients served by UFHCC, 37% of the population are residing in rural areas, and 33% of them are retired with a median age of 55 years. These patients need to make regular visits to manage their disease conditions and their appointments are typically made in advance. When visiting the medical center, in addition to the heavy travel burden, patients, especially elderly ones, have difficulty in locating the specialty clinic inside of the medical center, and the confusing check-in process could cause delays and disruptions in receiving care services. A service delivered to patient homes can reduce the negative patient experience in the medical center and expand access to care for patients with mobility barriers. To investigate this alternative service and its impact on care delivery system design, we introduce a game-theoretic model with the assumptions elaborated below.

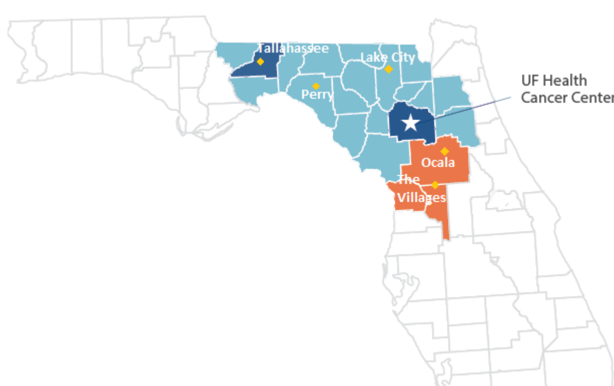


Figure 1 Catchment area of UFHCC (UF Health Cancer Center 2020).

Patient community We consider a set of communities (I) in the service region, characterized by two features: the average demand D_i and the average transportation cost f_i, incurred by patients traveling from community $i \in I$ to the central care facility. In the context of chronic disease or cancer management, in contrast to urgent or acute care, D_i represents the regular follow-up visits scheduled in advance, proportional to the size of the patient population. Therefore, we treat D_i as a deterministic variable. In addition, communities vary in their location and their proximity to the medical center. Notably, patients' mobility and disease conditions also contribute to their transportation costs. Without loss of generality, we encapsulate such heterogeneity using a unified measure of travel burden, denoted as f_i. The model can be generalized to investigate f_i as a function of proximity, disease burden, and mobility, among other factors.

Payment structure Patients pay a fixed amount of C_1 for the physician service (either in-person or virtual) (The Official U.S. Government Site for Medicare 2020). For VC visits, patients are also charged for the nursing service. If a discriminative pricing strategy is allowed, the medical institution can set the price for community i as $C_{2,i}$. If not, a uniform price $C_{2,f}$ is set. Patients pay a co-insurance $\eta_d C_{2,i}$ or $\eta_f C_{2,f}$ with $\eta_d, \eta_f \in [0,1]$, respectively. An alternative co-payment model suggests that patients pay a fixed C_2 to the medical institution for nursing services. This can represent the setting that the insurer (but not the medical institution) fully determines the patient's out-of-pocket payment. Since insured patients pay premiums to insurers regardless, we do not factor it into the utility for comparing VC and in-person visits.

The medical institution collects patients' out-of-pocket payments for both physician services and nursing services. For insured patients, the medical institution bills the insurers and gets a lump-sum payment. Other payers (e.g., governments) also subsidize or reimburse the medical institution for contracted care plans (e.g., Medicare and Medicaid). The reimbursement for physician services is thus omitted in the medical institution's revenue as payers typically reimburse physician services for VC and in-person visits at an equal rate (eVisit 2020), and the total patient demand stays the same. The reimbursement for auxiliary nursing services might also be a lump-sum payment by the insurer. The incumbent reimbursement policy does not cover nursing services for telehealth, and we consider a conservative scenario by omitting the insurer payment for nursing services for our major analysis. A discussion of non-zero reimbursement can be found in Section D in the Appendix.

Patient utility For patients, they gain $\mathcal{R}_k$, $k = 1, 2$, as the reward of receiving services. They also pay the service toll and assume the congestion and travel costs. The utilities of patients from community i for in-person and VC visits are defined as:

$$U_{i,1} = \mathcal{R}_1 - C_1 - f_i - \alpha \sum_{j \in I} \rho_j D_j, \quad \text{central care facility,} \tag{1}$$

$$U_{i,2} = \mathcal{R}_2 - C_1 - C_2 - \eta_d C_{2,i} - \eta_f C_{2,f}, \quad \text{community } i. \tag{2}$$

The term $\sum_{j\in I}\rho_j D_j$ represents the congestion (denoted in the following as W) in the central care facility. The congestion disutility for VC visits is not as pronounced as that in the in-person setting, where congestion implies a crowded environment, a higher risk of contracting an infectious disease (such as COVID-19), and a larger variability of services. Therefore, we only penalize this negative externality for in-person visits and use α to measure the sensitivity/unit effect of crowding on patients. For instance, a spike in telehealth demand was witnessed during the pandemic, where the adverse consequence of crowdedness measured by α is so large that patients' utilities of receiving in-person care are greatly reduced. Besides the risk of infection, patients generally favor a less crowded environment under which the per-patient space, staffing, and other resources are greater and the procedures are more ordered. Patients are strategic entities and they pick the option that maximizes their own utility, i.e., the choice i is obtained as $\arg\max_i\{U_{i,1}, U_{i,2}\}$. This forms the *incentive compatibility* (IC) constraints:

$$U_i = \max_i\{U_{i,1}, U_{i,2}\}, \quad \forall i \in I. \tag{3}$$

Remark: We keep $C_{2,i}$, $C_{2,f}$ and C_2 in the utility function for completeness, and η_d, η_f and C_2 will not be zero at the same time. In case the payment for provider services, C_1, is different between in-person and VC visits, the difference between them can be absorbed by C_2 mathematically as a payment adjustment.

Cost Structure of the Medical Institution First, the medical institution employs nurses to provide auxiliary care during the VC visits. Medical personnel such as nurses can help measure the vitals and conduct physical examinations that cannot be delivered virtually, and their training and knowledge can improve patients' compliance with the clinical guideline. These nurses can be supplied by the participating satellite clinics or home care service providers. Nurses are typically paid by a flat rate or under a contract (Schmidt 2020). The additional cost to the medical institution mainly comes from the productivity loss and the distance-based travel compensation. In particular, a nurse sees fewer patients when traveling to patient homes compared to staying at the clinic, and the travel cost depends on the number of homes to be visited. Thus, we consider a cost structure $g_i(D) = \beta_i D + \theta_i I_{D>0}$, i.e., the nurse coordination cost is proportional to the VC visit demand rate D plus a fixed set-up cost θ. We start with $g_i(D) = \beta_i D$, for which we provide an algorithm (Algorithm 1) that converges to the global optima. For a broader consideration of cost structures, we give an algorithm (Algorithm 2) that suits piece-wise linear concave functions as the approximation of general concave functions and is guaranteed to converge to the local optima. A sensitivity analysis regarding the cost function is presented in the numerical study. Other operational costs are invariant to telehealth services and hence are not factored into the model.

Second, the medical institution suffers a congestion penalty when the number of patients coming to the central care facility (e.g., a hospital) exceeds a threshold W_e. Congestion in the hospital could accelerate the spread of infectious disease (such as COVID-19). It could also cause the over-utilization of capacity-limited resources and affect the hospital's operating cost. For instance, when congestion presents in an understaffed setting, the hospital risks paying extra for agency nurses or paying at an overtime rate. Additionally, congestion attributes to medication errors and diagnostic delay and incurs other intangible costs. We penalize the medical institution for excessive congestion with a unit cost rate γ. Lastly, because the patient demand is stable and the physician supply is also unchanged, other operating costs are not significantly affected by VC visits and are not featured in the model.

Information structure We assume that patient demands, service prices, and congestion and transportation costs are common knowledge. Also, the medical institution is able to meet patient needs, i.e., the care provider is willing to cater to patients' needs. The setting of patients with strict preferences is discussed in Section 6.3. In addition, we consider a single major care provider in the service region, and the monopoly assumption can be relaxed by considering patient loss or patient migration, which is discussed in Section C in the Appendix. The variables and parameters introduced above are summarized in Table A.1 in the Appendix.

4. Discriminative Pricing Scheme

Nurses employed by nursing homes and assisted living facilities are often the onsite source through which physicians can assess patients using telehealth (Miller 2020). The cost of nursing service can be location dependent, which motivates the discriminative pricing strategy based on the patient's residency (Schmidt 2020). Without loss of generality, we start our analysis with the discriminative pricing scheme and provide the full course of analysis regarding the optimal pricing strategies to maximize the medical institution's revenue and the social welfare.

4.1. Revenue Maximization

Community hospitals are typically profit-driven. A monopolistic medical institution determines whether to offer VC visits at patient homes in different communities (option 2) to supplement in-person visits provided at the central care facility (option 1). It also determines the price $C_{2,i}$ for auxiliary nursing services provided in community i. Based on the prices, patients make decisions to maximize their perceived utility (the IC constraints) and they are indifferent to the type of care services received as long as the perceived utilities stay the same. To induce randomness, patients can employ a mixed strategy: in community i, a proportion $\rho_i \in [0,1]$ of patients receive in-person visits versus VC visits (1-ρ_i). The optimization problem (Model 1) is presented below:

$$\begin{aligned} \max_{C_{2,i}} \quad & \sum_i D_i(1-\rho_i)(\eta_d C_{2,i} + C_2) - \gamma \left[\sum_i \rho_i D_i - W_e\right]^+ - \sum_i g_i((1-\rho_i)D_i) \\ \text{s.t.} \quad & \text{IC constraints } (1)-(3). \end{aligned}$$

We let $\eta_d \neq 0$ and $\eta_f = 0$ to represent the discriminative pricing scheme, and the non-discriminative case ($\eta_d = 0$ and $\eta_f \neq 0$) is considered in Section 5. Here we claim that the medical institution can use pricing alone to navigate patients and the solution to Model 1 is the same as considering both $C_{2,i}$'s and ρ_i's as decision variables (see the Appendix for the proof). This equivalency greatly simplifies the analysis.

For a clear presentation of the main results, we start with the analysis of one and two communities that provides insights generalizable to multiple communities.

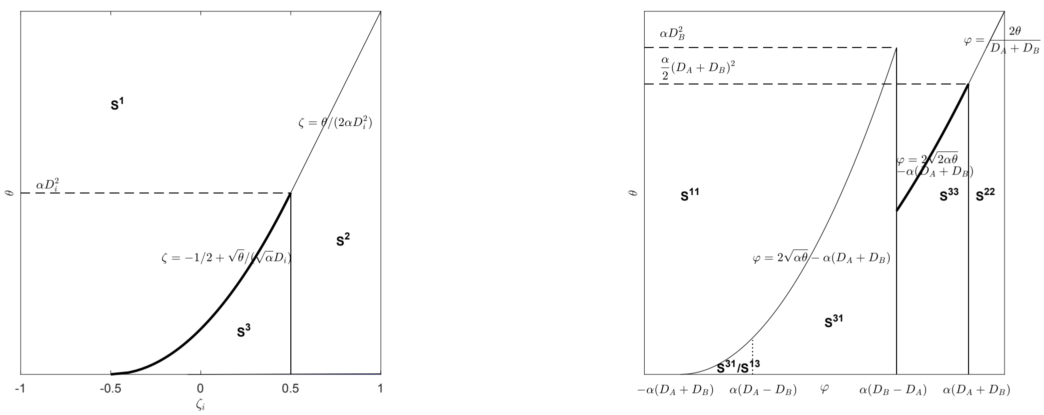
4.1.1. Single Community If the service region is rather uniform in demand and travel cost, it can be considered as one large community. For a single community (i), three scenarios (S) can unfold: (1) all patients go to the hospital, (2) all patients stay at home, and (3) part of them stay at home, denoted as S^m, $m \in \{1,2,3\}$. Let $\varphi_i = f_i - \beta_i + \gamma - \Delta\mathcal{R}$ represent the system's *marginal gain* from VC visits, where f_i is the saving of individual travel expenditure, β_i is the unit nurse coordination cost, γ is the medical institution's sensitivity to excessive congestion, and $\Delta\mathcal{R} = \mathcal{R}_1 - \mathcal{R}_2$. Here $\Delta\mathcal{R}$ represents the reward difference between the two services, which is positive when patients have a higher preference for in-person visits than VC visits, and negative, otherwise. For mathematical convenience, let $\zeta_i = \varphi_i/(2\alpha D_i)$. The equilibrium patient diversions responding to the optimal pricing are summarized in Proposition 1 and illustrated in Figure 2a.

Proposition 1. *With nurse cost function $g_i(D) = \beta_i D + \theta I_{D>0}$, the optimal patient diversions with one community are as follows:*

1. $\theta < \alpha D_i^2$. When $1/2 > \zeta_i > -1/2 + \sqrt{\theta}/(\sqrt{\alpha} D_i)$, a unique mixed strategy equilibrium exists, where $\rho_i = 1/2 - \zeta_i$ fraction of patients will go to the hospital and the rest stay at home. When $\zeta_i \geq 1/2$, or $\zeta_i \leq -1/2 + \sqrt{\theta}/(\sqrt{\alpha} D_i)$, all patients choose either home or hospital, respectively.

2. $\theta \geq \alpha D_i^2$. Only pure strategies exist, i.e., all patients stay at home when $\zeta_i \geq \theta/(2\alpha D_i^2)$ or go to the hospital, otherwise.

Proposition 1 suggests that when the set-up cost θ is high, a patient diversion is not favored: depending on the marginal gain from VC visits, all patients either choose VC visits or go to the hospital. With a small to moderate set-up cost, the percentage of in-person visits is decreasing



(a) Single community

(b) Two communities with an equal φ.

Figure 2 Parameter region partition examples.

with the marginal gain from VC visits. The boundaries of partitions are plotted in solid lines in Figure 2a. It manifests that a mixed strategy equilibrium (S^3) occurs when neither the set-up cost nor the absolute value of the marginal gain (or loss) is dominantly large. These results are consistent with our intuition.

Having $W_e \neq 0$ leads to a continuous but non-differentiable objective function and consequently complicates the optimization problem. However, it does not significantly impact the structure of optimal solutions, especially when γ is not very large. We defer the discussion of the impact of a positive congestion threshold W_e in the numerical study and more discussions can be found in the Appendix.

4.1.2. Two Communities With two communities A and B, a total of nine scenarios can occur, denoted as S^{mn}, $m, n \in \{1,2,3\}$. According to the uni-community analysis, the marginal utility change φ_i is a critical factor, and two cases, $\varphi_A = \varphi_B = \varphi$ and $\varphi_A \neq \varphi_B$ can unfold. We reserve the technical results in the Appendix and only present the major observations here.

Figure 2b illustrates the equilibrium patient strategies of communities A and B with an equal marginal gain in different parameter regions, under the assumption $D_A \leq D_B$. The case of $D_A \geq D_B$ is symmetric and can be analyzed analogously.

An important takeaway is that when the marginal utility gains are equal, there can be multiple equilibria. In particular, when $\theta = 0$ (no set-up cost), one can combine the two communities into one to significantly simplify the analysis. Since the unit utility gain is invariant to community membership, as long as the total fraction of patients going to the hospital $\bar{\rho} := (\rho_A D_A + \rho_B D_B)/(D_A + D_B) = 1/2 - \varphi/(2\alpha(D_A + D_B))$ is maintained, the optimal revenue can be achieved.

When $\varphi_A \neq \varphi_B$, different equilibrium patient strategies are exhibited in Figure 3. The partitions are generated with $D_A < D_B$ and are plotted on $\zeta_i = \varphi_i/(2\alpha D_i)$, $i = A, B$ for illustration purposes.

While the readers are referred to the Appendix for further technical details (Proposition A.2), we outline the main findings here.

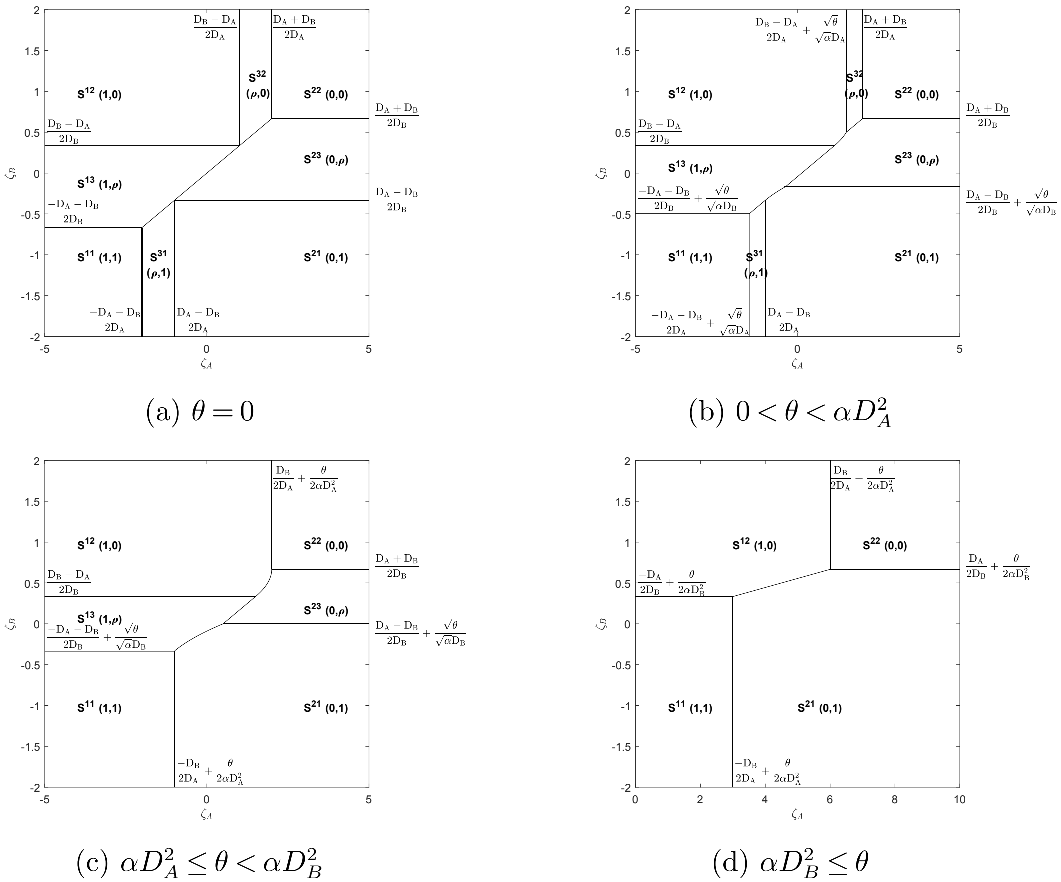
When the marginal gain or loss of community i is considerable, i.e., $|\varphi_i| > \alpha(D_A + D_B)$, its patients will stay at home or go to the hospital, regardless of the choice of the other community. When the marginal utility change of neither community is large enough, patients will consider others' behaviors, and the population size comes in to play. In addition, the two communities cannot take mixed strategies simultaneously, that is, S^{33} is not feasible. The medical institution desires patients with a larger φ_i to stay at home. As such, mixed strategy equilibria are elicited either when the difference in gains/losses between the choice of home and hospital is minimum, or patients' sensitivity to congestion α and/or the demand rate D_i is considerable. Both indicate that a minor change in the proportion of patients going to the hospital will have a salient impact on congestion and therefore nudging their preferences. In the case that patients are more sensitive to congestion than the medical institution, congestion becomes valuable to the medical institution — it desires patients to come to the hospital to generate a certain level of congestion so it can charge a higher price for the virtual service.

4.1.3. Multiple Communities The cost structure being $g_i(D) = \beta_i D + \theta_i I_{D>0}$ entails a knapsack problem, which is NP-hard. Therefore, we start with a linear nurse coordination cost $g_i(D) = \beta_i D$ and provide an $O(log(N))$ time algorithm. For general concave functions, we provide a well-performing heuristic solution to our problem.

Linear nurse coordination cost functions Without loss of generality, we assume there are N communities with unique φ_i's ranked in an ascending order. If there exist two communities $i, j \in I$ with $\varphi_i = \varphi_j$, then, combine them until getting unique φ_i's. This operation is justified based on the two-community analysis.

Theorem 1. *The optimal solution (ρ_i^*'s) to the revenue-maximizing problem with linear nurse coordination cost functions has the following structure. There exists a community $K \in I$ such that for $\varphi_i < \varphi_K$, $\rho_i^* = 1$, and for $\varphi_i > \varphi_K$, $\rho_i^* = 0$, and this community is nominated as the "threshold community." With K being the threshold community, if $\rho_K^* \in (0,1)$, the optimal congestion, i.e., the congestion generated by the optimal number of patients coming to hospital, is obtained as $W^* = \sum_i \rho_i^* D_i = (\alpha \sum_i D_i - \varphi_K)/(2\alpha)$.*

According to Theorem 1, at most one community (the threshold community) is optimal to adopt mixed strategies. This is consistent with the observation of two communities A and B with $\varphi_A \neq \varphi_B$. In addition, when $|\varphi_i|$ is large enough, community i patients will stay at home or go to the hospital, regardless of the choice of patients in other communities. This echoes the observation



(a) $\theta = 0$ (b) $0 < \theta < \alpha D_A^2$

(c) $\alpha D_A^2 \le \theta < \alpha D_B^2$ (d) $\alpha D_B^2 \le \theta$

Figure 3 Parameter region partitions for two communities with $D_B = 3D_A$ and $\varphi_A \neq \varphi_B$.

of the two-community case. In particular, when all φ_i's are large, it is optimal to let all patients stay at home (See Corollary A.1).

In short, we are delighted to find that the property of the system with multiple communities aligns with that of single and dual communities. In the following, we present the algorithm that offers us the exact solutions to the revenue-maximizing problem with linear nurse coordination cost functions. This is equivalent to find the threshold community K and ρ_K^*. To obtain K, we first calculate $W_k^C = \sum_{i<k} D_i$ and $H_k^C = \sum_{i>k} D_i$, and obtain $\rho_k^{<k>} = \max\{0, \min\{1, (D_k + H_k^C)/(2D_k) - (\varphi_k + \alpha W_k^C)/(2\alpha D_k)\}\}$, for each $k = 1...N$. Here $\rho_k^{<k>}$ is the optimal percentage of community k patients going to the hospital when community k is chosen as the "candidate threshold community," and the corresponding revenue is denoted as $R^{<k>}$. The true threshold community satisfies $K = \text{argmax}_k R^{<k>}$.

Proposition 2. *If $\rho_k^{<k>} \in (0,1)$, then, k is the true threshold community, i.e., $K = k$ and $R^{<k>}$ is the optimal revenue. If $\rho_k^{<k>} = 1$, then $R^{<k>} > R^{<k-1>}$; if $\rho_k^{<k>} = 0$, then $R^{<k>} > R^{<k+1>}$.*

Proposition 2 articulates that $R^{<k>}$ as a function of k is unimodal. Based on this property, we device Algorithm 1 (see Table 1).

Table 1 Algorithm 1 for linear nurse coordination cost functions.

	Initial $lower = 1, upper = N+1$.
For $k = \lfloor (lower + upper)/2 \rfloor$:	
	Calculate $W_k^C = \sum_{i<k} D_k$ and $H_k^C = \sum_{i>k} D_k$.
	Calculate the corresponding
	$\rho_k^{<k>} = \max\{0, \min\{1, (D_k + H_k^C)/(2D_k) - (\varphi_k + \alpha W_k^C)/(2\alpha D_k)\}\}$.
	If $\rho_k^{<k>} \in (0,1)$, $R^{<k>}$ is optimal, stop.
	If $\rho_k^{<k>} = 1$, $lower = k$, continue.
	If $\rho_k^{<k>} = 0$, $upper = k$, continue.
Until k does not change.	

Proposition 3. *Algorithm 1 for linear nurse coordination cost functions runs in $O(log(N))$ time, where N is the number of communities.*

This algorithm is computationally efficient — there is no need to evaluate the revenue in each iteration, and once a mixed strategy $\rho_k^{<k>}$ is obtained, the threshold community k is identified.

Concave nurse coordination cost functions First, we show that for general concave cost functions, there exists an optimal solution that dictates at most one community to adopt a mixed strategy. Notably, the notion of "general" here refers to a class of non-decreasing concave functions that cross the origin, conforming to the nature of cost functions. The cost structure being $g_i(D) = \beta_i D + \theta_i I_{D>0}$ is a special case in that class.

Theorem 2. *For a general concave nurse coordination cost function, there exists an optimal solution that dictates at most one community $k \in I$ to have $\rho_k \in (0,1)$, and all other communities have optimal patient diversions $\rho_i = 0$ or 1, $i \neq k$.*

For brevity, we introduce a vector $\boldsymbol{\rho}$ representing the patient diversion of all communities. With a general concave cost structure, the revenue maximization problem is NP-hard and the optimal solution $\boldsymbol{\rho}^{\text{gen}}$ is difficult to obtain. However, Theorem 2 elucidates a nice structure, which reveals that the structure of optimal solutions to the problem with general concave cost functions can be similar to that of linear ones. It motivates us to approximate $\boldsymbol{\rho}^{\text{gen}}$ using a linear approximation. In the following, we provide a theorem showing that the optimal solution $\boldsymbol{\rho}^{\text{lin}}$ solved under the linear cost structure assumption is a good approximation.

Theorem 3. *The optimal solution $\boldsymbol{\rho}^{lin}$ under the linear cost structure with $g_i^{lin}(D) = \beta_i D$, $\beta_i = g_i(D_i)/D_i$ is a feasible solution to the revenue maximization problem under the general concave cost structure $g_i(D)$, and the corresponding revenue $R(\boldsymbol{\rho}^{lin})$ approximates the optimal objective function $R(\boldsymbol{\rho}^{gen})$ with a gap less than $g_K((1-\rho_K)D_K) - \beta_K(1-\rho_K)D_K$, where K is the threshold community in $\boldsymbol{\rho}^{lin}$.*

The constraints of the revenue-maximizing problem remain invariant to the nurse coordination cost function, which only appears in the objective function, and hence, any feasible solution to the linear problem is also feasible to the concave one. The choice of the linear approximation $g_i^{\text{lin}}(D) = \beta_i D$ is motivated by the fact that $g_i^{\text{lin}}(D_i) = g_i(D_i)$ and $g_i^{\text{lin}}(0) = g_i(0) = 0$. When communities have optimal patient diversions $\rho_i = 0$ or 1, this approximation is exact. This theorem is very powerful because a good approximation is only required for the threshold community but not others, and if the cost function for the threshold community is indeed linear, then, no matter how other cost functions look like, the optimal revenue is guaranteed, i.e., $R(\boldsymbol{\rho}^{\text{gen}}) = R(\boldsymbol{\rho}^{\text{lin}})$.

Without solving the optimization problem, if we know that $\max_{i\in I} g_i(D_i)$ is small in scale compared to revenue $R(\cdot)$, we can directly use the solution obtained under the linear assumption to approximate that under the concave assumption. Only when $g_K((1-\rho_K)D_K) - \beta_K(1-\rho_K)D_K$ is not negligible, we then need to consider a better approximation of $g_i(D)$, such as a piece-wise linear concave function. Consequently, we introduce Algorithm 2, where the optimization problem turns into a non-linear integer program.

Assume there are $L_i \geq 1$ segments for each piece-wise linear concave function $g_i^{\text{p-lin}}(D)$ as an approximation to $g_i(D)$, and let β_i^l be the gradient of the cost function at the l^{th} segment, $1 \leq l \leq L_i$. Let $\varphi_i^l = f_i + \gamma - \beta_i^l - \Delta\mathcal{R}$ for $1 \leq l \leq L_i$. In addition, introduce $\rho_i^{t,l}$ as the optimal diversion of community i at iteration t by assuming the gradient of the cost function falls on the l^{th} segment with value β_i^l. Define $R^{t,l} = R(\rho_i^{t,l})$ as the optimal revenue when community i adopts the optimal solution $\rho_i^{t,l}$ and the other communities keep their current diversions, i.e., ρ_j^t for $j < i$ and ρ_j^{t-1} for $j > i$ (the definitions are provided in Table 2). We deem a solution $\rho_i^{t,l}$ as feasible if $(1-\rho_i^{t,l})D_i$ falls on the domain of segment l; otherwise, it cannot serve as a candidate for the optimal solution. Furthermore, for the cost function with a non-zero set-up cost, denote $\rho_i^t = 1$ (implying no nurse coordination cost) as $\rho_i^{t,l=0}$, and the corresponding revenue as $R^{t,l=0}$. Then, we define the optimal solution of ρ_i at iteration t as $\rho_i^t = \operatorname{argmax}_{\rho_i^{t,l}} R^{t,l}$, for $0 \leq l \leq L_i$. $H_{t,i}$ ($W_{t,i}$) is the amount of stay-home (go-to-hospital, respectively) patients at iteration t after updating patient diversion ρ_i^t at the i-th community.

Proposition 4. *Algorithm 2 for piece-wise linear concave nurse coordination cost functions converges to a local optimum.*

This algorithm is a special case of the minorize-maximization algorithm, which is an iterative optimization method that finds the maxima of convex programs. Since this problem entails a non-linear integer program, the algorithm converges to a local optimal in general. The proof of convergence can be found in the Appendix.

Table 2 Algorithm 2 for piece-wise linear concave nurse coordination cost functions.

Initialization with the optimal solution $\boldsymbol{\rho}^{lin}$.
For t=1...
For i =1...N:
If i=1, $H^C_{t,i} = H_{t-1,N} - (1-\rho_i^{t-1})D_i$, $W^C_{t,i} = W_{t-1,N} - \rho_i^{t-1}D_i$.
If i=2...N, $H^C_{t,i} = H_{t,i-1} - (1-\rho_i^{t-1})D_i$, $W^C_{t,i} = W_{t,i-1} - \rho_i^{t-1}D_i$.
For each segment $0 \le l \le L_i$:
Calculate $\rho_i^{t,l} = \max\{0, \min\{1, (D_i + H^C_{t,i})/(2D_i) - (\varphi_i^l + \alpha W^C_{t,i})/(2\alpha D_i)\}\}$.
Check feasibility, if feasible, calculate $R(\rho_i^{t,l})$.
Update $\rho_i^t = \operatorname{argmax}_{\rho_i^{t,l}} R^{t,l}$.
Update $H_{t,i} = H^C_{t,i} + (1-\rho_i^t)D_i$, $W_{t,i} = W^C_{t,i} + \rho_i^t D_i$.
Until converge.

4.2. Value of VC visits

We show here that under certain conditions, the medical institution would like to offer VC visits, which strictly benefits both the medical institution and patients.

PROPOSITION 5. *With linear nurse coordination cost functions, when* $\varphi_{max} = \max_{i\in I}\varphi_i > -\alpha\sum_{j\in I} D_j$, *some patients prefer to receive care at home, and the medical institution and patients who remain to go to the hospital are strictly better off. Moreover, the medical institution collects strictly more money from every patient who switches from going to the hospital to staying at home.*

Proposition 5 suggests that the largest marginal utility gain from offering VC visits should exceed a threshold $-\alpha\sum_i D_i$. In practice, if the overall cost to deploy VC visits is large due to a high set-up cost (cost irrelevant to unit demand, e.g., equipment fee or bonus paid to nurses who are on travel) and/or a high unit cost for dispatching nurses to patient homes, the medical institution would not implement VC visits.

Furthermore, if there exist patients receiving VC visits, then, the medical institution collects strictly more money with the amount of $\varphi_i + \alpha W^* > 0$ from every at-home patient in community i, where W^* is the optimal congestion level. Notably, not all regions are equally profitable as a result of patient heterogeneity. For instance, if all nursing service costs are identical, the medical institution favors patients from larger-scale communities that are far from its central care facility to receive VC visits, so as to reap strictly more surplus from patients.

4.3. Social Welfare Maximization

In contrast to the medical institution, a social planner aims to maximize the social welfare with the following optimization problem under the linear nurse coordination cost assumption:

$$\begin{aligned}\max_{C_{2,i}} \quad & \sum_{i\in I}(1-\rho_i)D_i[\varphi_i + \alpha\sum_{j\in I}(1+\rho_j)D_j] \\ \text{s.t.} \quad & \text{IC constraints } (1)-(3).\end{aligned}$$

PROPOSITION 6. *For social welfare maximization under the linear nurse coordination cost assumption: If $\varphi_{max} > -2\alpha\sum_{j\in I} D_j$, then, there are patients who should receive care at home, and both the aggregate patient surplus and the social welfare are strictly larger.*

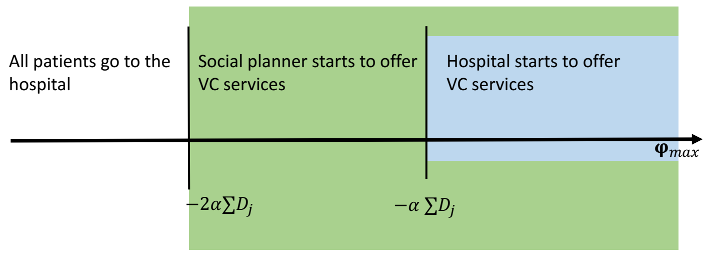


Figure 4 Conditions for implementation of VC visits

Figure 4 compares the feasible regions for the social planner and the medical institution to offer VC visits. If the maximum marginal gain φ_{max} is smaller than $-2\alpha\sum_{i\in I} D_i$, then both of them would not offer VC visits. When φ_{max} increases to $(-2\alpha\sum_{i\in I} D_i, -\alpha\sum_{i\in I} D_i]$, the social planner starts to ask some patients to stay at home while the medical institution still sends them all to the central hospital, according to Proposition 5. When φ_{max} continues to increase, offering VC visits becomes a consensus, but the congestion generated by the social welfare maximizer is always less than that of the revenue maximizer, as elaborated in Proposition 7 below.

We introduce superscript "Wel" to refer to the results obtained from social welfare maximization. Let ρ_i^{Wel} be the optimal patient diversion of community i, and $W^{\mathrm{Wel}} = \sum_i \rho_i^{\mathrm{Wel}} D_i$ be the optimal amount of patients going to the hospital. We further introduce superscript "Rev" to refer to the results obtained from revenue maximization. Let $\Delta_{\mathrm{Hospital}}$, $\Delta_{\mathrm{Patients}}$, and $\Delta_{\mathrm{Welfare}}$ be the difference in revenues between with and without offering virtual services, and that in patient surplus and social welfare, respectively.

PROPOSITION 7. *Under the conditions that $\varphi_{max} > -2\alpha\sum_{j\in I} D_j$ and $\exists\ i$, $\varphi_i < 0$, comparing revenue maximization and social welfare maximization:*

1. *The optimal congestion based on the social planner's decision W^{Wel} is strictly less than that from the medical institution's choice W^{Rev}.*
2. *$\forall i \in I$, $\rho_i^{Wel} \le \rho_i^{Rev}$ and there exists at least one $i \in I$ such that $\rho_i^{Wel} < \rho_i^{Rev}$.*
3. *$\Delta^{Wel}_{Patients} > \Delta^{Rev}_{Patients}$, $\Delta^{Wel}_{Hospital} < \Delta^{Rev}_{Hospital}$, and $\Delta^{Wel}_{Welfare} > \Delta^{Rev}_{Welfare}$.*

It can be observed that the medical institution prefers more patients coming to the hospital than the social planner does. By maintaining a certain level of "crowdedness," the medical institution can charge a higher price for at-home patients and gain more revenue. In addition, unlike patients who are no worse under the social welfare maximization, the increase in the medical institution's revenue is not guaranteed. While the "artificial congestion" is not favored by the social planner, the alignment between the social planner and the medical institution can be established and the discussions are presented in Section 6.1.

5. Non-discriminative Pricing Scheme

In this section, community-specific pricing for VC visits is not allowed ($\eta_d = 0$). Specifically, two cases can unfold: (i) $\eta_f = 0$ and patients pay a constant C_2 that the medical institution can not determine (for instance, determined by the insurer); (ii) $\eta_f \neq 0$, and a sum of $\eta_f C_{2,f} + C_2$ is paid for nursing services. For both cases, we obtained the following patient strategies.

THEOREM 4. *When patients from all communities pay the same for VC visits:*

(a) (Monotonicity in travel burden) For any $f_i > f_j$, if community j patients are optimal to stay at home, then community i patients are also optimal to stay at home.

(b) (Threshold structure) There exists a threshold travel burden $F = \eta_f C_{2,f} + C_2 - \alpha W + \Delta\mathcal{R}$, where $W := \sum_{i\in I} \rho_i D_i$ is the total number of patients coming to the hospital. For community i patients, if $f_i > F$, stay at home; if $f_i < F$, go to the hospital; if $f_i = F$, a mixed strategy equilibrium can exist.

This theorem shows that with a flat VC visit payment, the patient strategy still enjoys a threshold policy. The travel burden plays an essential role in this case. Instead of ranking the communities according to their marginal gain φ_i's to identify the threshold community, now the ranking is based on their travel burdens and one aims to find out the threshold travel burden.

In case (i), the medical institution cannot negotiate the price with the insurer due to a lack of market power. Then, the medical institution is willing to offer VC services only when the price C_2 is greater than a threshold as described in Eq. (A.8) (see Proposition A.4 in the Appendix).

In Case (ii), the medical institution aims to find the optimal solution to the following problem:

$$
\begin{aligned}
\max_{C_{2,f}} \quad & (\eta_f C_{2,f} + C_2)\left(\sum_{j:f_j>F} D_j + (1-\rho_i)D_i\right) - \sum_{j:f_j>F} g_j(D_j) - \sum_{j:f_j=F} g_j((1-\rho_j)D_j) \\
& -\gamma\left[\sum_{j:f_j<F} D_j + \sum_{i:f_i=F} \rho_i D_i - W_e\right]^+ + \gamma\left[\sum_{j\in I} D_j - W_e\right]^+ \\
\text{s.t.} \quad & \text{IC constraints (1)} - \text{(3)}.
\end{aligned}
$$

According to Theorem 4, only the threshold community can adopt a mixed strategy, denoted as ρ_i. Then, we claim that the above problem is the same as the one considering both $C_{2,f}$ and ρ_i as decision variables (see the Appendix for the proof).

We show here that under certain conditions, the medical institution would like to offer VC visits, which strictly benefits both the medical institution and patients.

PROPOSITION 8. *With linear nurse coordination cost functions, when there exists a community i which has $\varphi_i > \alpha\left(\sum_{j\in I:f_j>f_i} D_j - \sum_{j\in I:f_j\leq f_i} D_j\right)$, some patients will prefer to receive care at home, and the medical institution and patients who remain going to the hospital are strictly better off.*

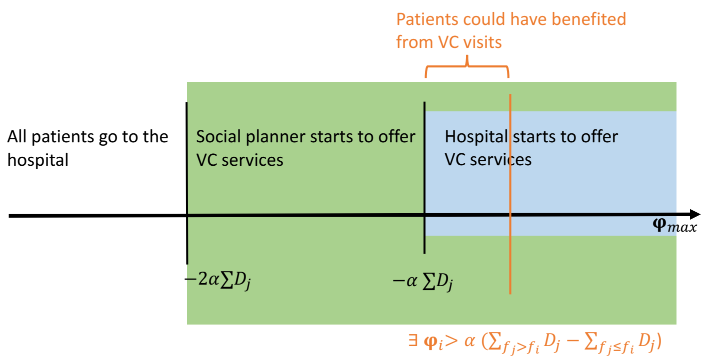


Figure 5 Conditions for implementation of VC visits under flat nursing service price

Since $\varphi_{max} = \max_{i\in I}\varphi_i \geq \varphi_i > \alpha(\sum_{f_j>f_i} D_j - \sum_{f_j\leq f_i} D_j) \geq -\alpha\sum_{j\in I} D_j$, the condition in Proposition 8 is stricter than that in Proposition 5. As shown in Figure 5, when discriminative pricing is not allowed, it is more difficult for the medical institution to benefit from offering VC visits and subsequently implementing it. This can actually hurt patients at the mean time. For instance, if $\max_{i\in I}\varphi_i > -\alpha\sum_{j\in I} D_j$, but for all i, $\varphi_i < \alpha(\sum_{f_j>f_i} D_j - \sum_{f_j\leq f_i} D_j)$, then, the medical institution will not offer the new service under the flat rate requirement. Nonetheless, it is optimal to do so if discretionary prices are allowed. According to Proposition 5, patients could have benefited from VC visits in this region, but they will miss this opportunity because the medical institution will not implement it.

One might expect that charging the same price brings "fairness"; however, this "fairness" can bring unintended consequences. When discriminative pricing is allowed, the system-level marginal gain φ_i is used to navigate patients. To achieve a larger marginal gain from VC visits, in addition to being far away from the central hospital, a community can also take advantage of the low nurse coordination cost to qualify for VC visits. In contrast, with a flat price, the threshold is solely based on travel burden f_i. A community with a larger f_i comparing to its peers is more likely to receive VC visits. However, choosing communities in this way inevitably include communities whose nurse coordination costs are actually large, which is not good for the overall system efficiency. To compensate the loss due to large nursing service costs, the medical institution will charge an overall higher price to every community. For the communities that enjoy a low nurse coordination cost, they could have been charged $C_{2,i}$ such that $\eta_d C_{2,i} < \eta_f C_{2,f}$, if a community-dependent price is allowed. Examples of such communities that are disproportionally affected by the fairness requirement are further illustrated in Section 7.3.

The social welfare gain from the new service compared to that without VC visits can be determined as:

$$\max_{C_{2,f}} \sum_i D_i(1-\rho_i)(\eta_f C_{2,f} + C_2) - \sum_i g_i((1-\rho_i)D_i) - \gamma \left[\sum_i \rho_i D_i - W_e\right]^+ + \gamma \left[\sum_{j\in I} D_j - W_e\right]^+$$
$$+ \sum_i \rho_i D_i \left(\alpha \sum_j (1-\rho_j) D_j\right) + \sum_i (1-\rho_i) D_i (-\Delta\mathcal{R} - (\eta_f C_{2,f} + C_2) + f_i + \alpha \sum_j D_j)$$
$$\text{s.t. IC constraints } (1)-(3).$$

PROPOSITION 9. *With linear nurse coordination cost functions, when there exists a community i which has $\varphi_i > -2\alpha \sum_{f_j \le f_i} D_j$, some patients prefer to receive care at home, and the social welfare is strictly better off.*

The comparison between the choices of the social planner and the medical institution under non-discriminative pricing is similar to that without this constraint. The condition $\exists\, \varphi_i > -2\alpha \sum_{f_j \le f_i} D_j$ is less restrictive than that for the medical institution to start VC visits. Moreover, since $\varphi_{max} \ge \varphi_i > -2\alpha \sum_{f_j \le f_i} D_j \ge -2\alpha \sum_{j\in I} D_j$, the condition for the social planner to implement VC visits under non-discriminative pricing is stricter compared with that for the discriminative case (see Proposition 6).

6. Discussions

6.1. Alignment between the Medical Institution and the Social Planner

While the social welfare maximizer brings no harm to patients, it is unclear if it is acceptable by the medical institution, and we present in Corollary 1 about the conditions that align the interests of the social planner and the medical institution under the discriminative pricing setting.

COROLLARY 1. If $\varphi_{max} > -2\alpha \sum_i D_i$, the social planner will offer VC visits, and the revenue gain or loss of the medical institution depends on the pricing strategy for VC visits employed by the social planner.

- A setting that favors a higher price of $C_{2,i}^{\mathrm{Wel}}$, such as using the upper bound of equilibrium prices, leads to an increased revenue (compared to no virtual service).
- An unfavorable price, such as the lower bound of equilibrium prices, leads to a reduced revenue (strictly lower than no virtual service, except for two special cases when $W^{\mathrm{Wel}} = 0$ or $W^{\mathrm{Wel}} = \sum_{j\in I} D_j$).

The equilibrium pricing: for the communities that satisfy $-\alpha W^{\text{Wel}} \leq \varphi_i$, $-C_2 - \Delta\mathcal{R} - \gamma + \beta_i \leq \eta_d C_{2,i}^{\text{Wel}} \leq -C_2 - \Delta\mathcal{R} + f_i + \alpha W^{\text{Wel}}$; for other communities, $\eta_d C_{2,i}^{\text{Wel}} = -C_2 - \Delta\mathcal{R} + f_i + \alpha W^{\text{Wel}}$.

This corollary shows that the optimal pricing of the social welfare maximization problem is not unique. Because the price is an internal transaction and does not feature in the objective function, we find a proper setting of prices, $\eta_d C_{2,i}^{\text{Wel}} = -C_2 - \Delta\mathcal{R} + f_i + \alpha W^{\text{Wel}}$, such that when some patients are willing to receive care at home, all three parties (patients, the medical institution, and the social planner) are strictly better off. This creates a first-best solution that is naturally incentive compatible for the medical institution. Notably, unlike the revenue maximizer, where the medical institution earns no less money from every patient, now it receives less money from each community j with $\varphi_j < -\alpha W^{\text{Wel}}$. Such communities exist when $W^{\text{Wel}} > 0$.

The pricing strategy that guarantees an improvement of all three parties (patients, the medical institution, and the social planner) only exists in the discriminatory pricing setting. If not allowed, there could still be a range of equilibrium prices, but even the one that is most favorable by the medical institution (i.e., the upper bound of $C_{2,f}$) is taken, the profitability of the medical institution is not guaranteed. An illustrative example can be found in Section 7.3.

6.2. Service Region Expansion

In our study, a patient in community i would seek care with the medical institution if $\max\{U_{i,1}, U_{i,2}\} \geq 0$. For an implementability analysis, the patient demand D_i represents the equilibrium population before the implementation of VC visits such that $U_{i,1} = \mathcal{R}_1 - C_1 - f_i - \alpha \sum_{j\in I} D_j \geq 0$. In this case, except for the community with the largest f_i, all other communities have a strictly positive utility, which means the demand rate D_i has already exploited all the potential demands in that community. It is not an equilibrium if there are unmet demands in community i but patients from another community with $f_j > f_i$ are catered for.

Let $D_{\text{All}} = \sum_{i\in I} D_i$ be the congestion in the hospital without the VC visit option, and W^* is the optimal congestion in the hospital where discriminative pricing is allowed. When the VC visit is implementable, W^* is strictly smaller than D_{All}. Here we consider the new catchment area where patients still have the option to go to the hospital, i.e., $U_{i,1} \geq 0$, when the congestion level W at the central hospital drops after the implementation of VC visits. With an expanded service region, the new community set is denoted as $\tilde{I}$, and $I \subset \tilde{I}$.

The expansion in service region introduces more *potential* demands. The medical institution might not be able to absorb these patients because of the capacity cap, e.g., limited number of physicians. However, it can still benefit from the service region expansion. The new optimization problem adds one more set of decision variables — the communities in the service region that the medical institution aims to cover owing to the capacity cap. Clearly, $\tilde{\Delta}^{\text{Rev}}_{\text{Hospital}} \geq \Delta^{\text{Rev}}_{\text{Hospital}}$, due to an expanded search space. Furthermore, we observe that, if the medical institution can reject or select patients in its catchment area to maximize its revenue, it desires the communities with greater φ_i's and larger demand D_i's. Due to limited provider capacity, serving the communities close to the hospital then becomes not as preferable, especially those with a small D_i or a small marginal benefit φ_i. To see this, consider an example where $I = \{1\}$, and $\tilde{I} = \{1, 2\}$, with $f_1 < f_2$. With limited service capacity, the medical institution can only pick one community to serve. Previously, when $-\alpha D_1 \leq \varphi_1 \leq \alpha D_1$, the maximum revenue is $R^1 = (\alpha D_1 + \varphi_1)/(4\alpha) - \theta$ according to a single community analysis. However, when $\alpha D_1 + \varphi_1 < \alpha D_2 + \varphi_2$, the medical institution is actually better off with community 2 as their only client with $R^2 = (\alpha D_2 + \varphi_2)/(4\alpha) - \theta$. And $R^2 > R^1$ when $D_2 > D_1$ (small demand) and/or $\varphi_2 > \varphi_1$ (low marginal benefit). In addition, if we restrict the total number of patients a medical institution can serve, it would rather serve one large community than multiple smaller communities with the same total number of patients (demands), especially when the set-up cost is high.

We further consider that the number of patients each physician sees can be elastic to a certain extent. This is evidenced by the fact that VC visits can reduce variabilities and provider non-value added time. Then, with the service region expansion, more patients can be served by the medical institution. Let $\tilde{W}^*$ be the optimal congestion with the expanded catchment area. As a result, the optimal congestion with the service region expansion should be strictly less than the congestion when no virtual service is provided, i.e., $\tilde{W}^* < D_{\text{All}}$. However, whether existing patients will be harmed by the expansion remains to be explored. For them, whether W^* is greater or less than $\tilde{W}^*$ is critical. A low congestion level in the hospital would curb the price of at-home care due to patients' IC constraints, i.e., each patient is willing to pay a lower price when $\tilde{W}^* < W^*$. They strictly benefit from less congestion and a lower price.

The choice of $\tilde{W}^*$ will be determined by the medical institution. On the one hand, pushing $\tilde{W}^*$ to be as small as possible will attract more communities (with a larger f_i) in the service region. The medical institution loses revenues from the existing patients but is compensated more from newly attracted patients. On the other hand, as the catchment area increases, more patients would *potentially* go to the hospital. The medical institution can encourage them to create a moderate hospital congestion level and reap the surplus by charging a higher price for at-home care. In this case, the existing patients would all be worse off.

6.3. Patient Segmentation

The optimal pricing strategy implies that at most one community can have a mix of in-person and VC visits. However, this is rarely seen in reality as a segment of patients might have fixed preferences. Therefore, we consider an extension of our basic discriminative pricing model with the inclusion of patients who strictly prefer in-person visits to VC visits and vice versa, in each community. We assume a certain portion of patients prefer to go to the hospital no matter how severe the congestion in the hospital is, and denote the total number as W^{Fix}. There are also patients who strictly prefer VC visits if it is offered, with the amount h_i^{Fix} for community i and a total of H^{Fix}. The rest population is denoted as D_i^{Flex} as before, and the partition $\rho_i \in [0,1]$ is calculated based on the flexible population D_i^{Flex}.

With the two pre-determined groups, the patients' utility staying at home remains the same and that going to the central care facility is updated as:

$$U_{i,1} = \mathcal{R}_1 - C_1 - f_i - \alpha \sum_{j\in I} \rho_j D_j^{\text{Flex}} - \alpha W^{\text{Fix}}, \quad \text{central care facility.}$$

If some patients strictly prefer VC visits regardless of prices, the medical institution may set the price arbitrarily high to maximize their revenue, especially if these patients dominate the total population. Therefore, we assume these patients strictly prefer VC visits when facing the same utility, and use the IC equilibrium price to calculate the total revenue.

If the medical institution only offers VC visits for the VC service "advocates," the total revenue is calculated as

$$R^{\text{All}} = \max_{C_{2,i}} \sum_i h_i^{\text{Fix}}(\eta_d C_{2,i} + C_2) - \gamma \left[\sum_i D_i^{\text{Flex}} + W^{\text{Fix}} - W_e\right]^+ - \sum_i \beta_i h_i^{\text{Fix}}.$$

Under the linear assumption, we have

$$R^{\text{All}} = \sum_{i\in I} h_i^{\text{Fix}}(\varphi_i + \alpha \sum_{j\in I} D_j^{\text{Flex}} + \alpha W^{\text{Fix}}).$$

The decision of providing VC visits exclusively for this segment of patients depends on whether the fees collected from them cover the associated nurse costs. When $R^{\text{All}} > 0$, the hospital would like to provide VC visits for these advocates, and a necessary condition is $\varphi_{max} > -\alpha \sum_{j\in I} D_j^{\text{Flex}} - \alpha W^{\text{Fix}}$. Since $\sum_{j\in I} D_j^{\text{Flex}} + W^{\text{Fix}} < \sum_{j\in I} D_j$ when $H^{\text{Fix}} > 0$, according to Proposition 5, the medical institution is actually less in favor of offering VC visits if more patients strictly prefer VC visits.

Next, we consider the situation that the medical institution not only provides VC services for the VC service advocates, but also the flexible population. Then, the total revenue becomes:

$$\begin{aligned} R^{\text{All}} = \max_{C_{2,i}} &\sum_i \left(h_i^{\text{Fix}} + D_i^{\text{Flex}}(1-\rho_i)\right)(\eta_d C_{2,i} + C_2) - \gamma \left[\sum_i \rho_i D_i^{\text{Flex}} + W^{\text{Fix}} - W_e\right]^+ \\ &- \sum_i \beta_i((1-\rho_i) D_i^{\text{Flex}} + h_i^{\text{Fix}}). \end{aligned}$$

Applying the same methodology as in the main analysis, we find that the threshold structure still exists, but a modified "optimal" congestion is applied: $W^{\text{Flex},*} = (\alpha \sum_i D_i^{\text{Flex}} - \varphi_K - \alpha W^{\text{Fix}} + \alpha H^{\text{Fix}})/(2\alpha)$. Moreover, the implementability condition becomes $\varphi_{max} = \max_{i \in I} \varphi_i > -\alpha \sum_{j \in I} D_j^{\text{Flex}} - \alpha W^{\text{Fix}} + \alpha H^{\text{Fix}}$, for which, the medical institution is willing to provide VC services for the flexible patients (and naturally the advocates will receive VC services). We can see that with more patients strictly favoring VC services, i.e., a larger H^{Fix}, the implementability condition becomes more strict. On the contrary, with more patients pre-determined to go the hospital (a larger W^{Fix}), the medical institution is more likely to offer VC visits.

One might expect the medical institution to be in favor of offering VC visits since there are patients strictly prefer VC visits. However, we have shown that this is not the case. Without patients having fixed preferences, the medical institution can "select" the most profitable population to offer VC visits; however, with the advocates, the medical institution has to cater to their demands first. Also, the medical institution would prefer a higher congestion, but its power to create "artificial congestion" is impaired due to the strict preference of the VC service advocates.

7. Numerical Study

7.1. Model Calibration

We use UFHCC as a motivating example and provide a model and investigate their VC visit service design. Unlike primary care, specialty care clinics and specialists (e.g., oncologists) for chronic disease and cancer management are typically located in urban areas or city centers, and patients might need to travel a long distance to access care. Currently, UF Health telehealth provides VC visits for their patients with home internet access, and we further assume those patients can receive telehealth at home with the presence of nurses from community-based outpatient clinics or home care providers. We examined the cancer registry data from UFHCC in calendar year 2018 to understand the service region it covers and the demand distribution. Newly registered cancer patients, especially those who just received surgeries need frequent follow-up visits by specialists to manage their cancer care, where telehealth suits well in this context. We analyzed the number of newly registered cancer patients by city and county (zip code indexed) to estimate the potential patient demands D_i across regions. The travel burden of patients to go to the nearest UF Health clinic available in their own county is used to estimate the nurse coordination cost rate β_i. The distance between patient homes and the medical center is used to estimate the individual travel burden f_i. Demand rate, travel cost, nurse coordination cost, and other parameters are normalized for illustration purposes. The parameter settings can be found in Tables A.7 and A.8 in the Appendix.

7.2. Results under the Discriminative Pricing Setting

A total of 3481 patients from 22 counties are considered in the numerical study. The most distant county covered is Gadsden County, FL, which is 190 miles away, about a three-hour drive to the medical center. The revenue-maximizing policy finds the Marion county (near Ocala, FL, see Figure 1) as the threshold community. Among the 22 counties, Marion county is ranked the 8^{th} based on distance, which is about a 40 minutes' one-way drive to the medical center. Marion has the largest demand size and is evaluated as the 12^{th} least profitable county. According to our policy, all counties that have a smaller marginal profit than Marion and 43.7% of the patients in Marion should go to the medical center, and the rest should stay at home, which accounts for 50.5% of the whole normalized demand (the relative market size of telehealth).

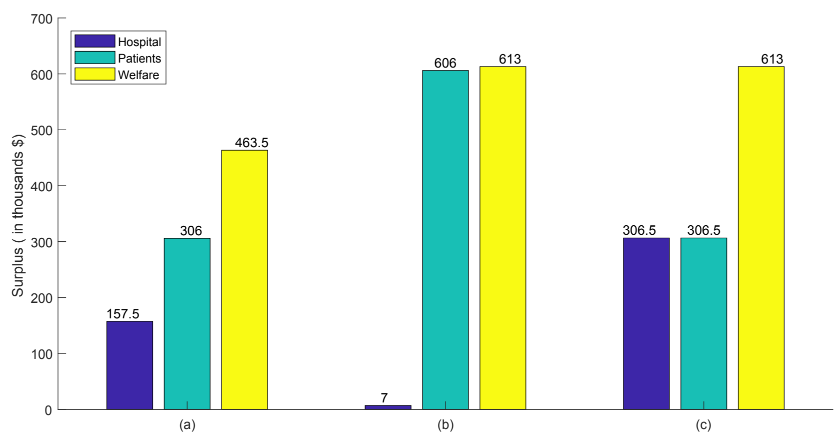


Figure 6 Benefits brought by telehealth for hospital revenue, patient surplus, and social welfare under different settings: (a) revenue maximization, (b) social welfare maximization, and (c) social welfare maximization with increased in-person care price C_1

Alignment of Revenue and Social Welfare Maximization As pointed out in Proposition 7, it may not be the best interests of the medical institution to fully comply with the social welfare maximizer as it can benefit strictly more from the revenue maximizer. In the following, we investigate a variation of the problem that could improve social welfare without compromising the medical institution's profitability. We compare the optimal solutions under (a) revenue maximization, (b) social welfare maximization, and (c) social welfare maximization with increased in-person care price C_1. The change in hospital revenue, patient surplus, and social welfare are exhibited in Figure 6, by assuming \$0.05 worth for each mile based on 3481 patients registered per year.

In baseline parameter settings, the medical institution and its patients are strictly better off compared with no telehealth. The revenue-maximizing policy favors the medical institution while the social-welfare maximizer favors patients, where the medical institution is barely more profitable. The medical institution's revenue drops significantly owing to the loss of "artificial congestion". In fact, all patients are optimal to stay at home under social-welfare maximization (based on Corollary A.2).

To increase social welfare while being incentive compatible with the medical institution, one can set the in-person payment C_1 to $C_1^M = C_1 + \alpha(W^{\text{Rev}} - W^{\text{Wel}})$ and decrease the adjustment payment for VC visits, C_2, to $C_2^M = C_2 - \alpha(W^{\text{Rev}} - W^{\text{Wel}})$. Notably, C_1 and C_2 are not decision variables but input parameters in the optimization problems. As a result, both the social planner and the medical institution are willing to set $\eta_d C_{2,i}^M = -C_2 - \Delta\mathcal{R} + f_i + \alpha W^{\text{Rev}} = -C_2^M - \Delta\mathcal{R} + f_i + \alpha W^{\text{Wel}}$ as the optimal pricing strategy. The results are presented in Scenario (c) in Figure 6. Under the social welfare maximizer, less patients are optimal to go to the hospital so the congestion at the hospital is alleviated. As such, it is not naturally incentive compatible for patients to accept the price of telehealth charged by the medical institution under revenue maximization. Therefore, external forces are needed to channel patients. For instance, the medical institution can raise patients' awareness of the intangible benefits of telehealth. The most frequently cited factors for patient satisfaction are convenience and reduced travel cost (Bohnenkamp et al. 2004), reduced wait time for the appointment and consultation, effective communication with the care provider (Mair et al. 2000, Bohnenkamp et al. 2004, Laila et al. 2008), and overall ease of use and quality of picture and sound (Laila et al. 2008). In addition, since the overall congestion at the hospital is reduced thanks to telehealth, the corresponding surplus released to patients can be further reaped by the medical institution through charging a higher in-person visit price. This can be justified as the overcrowding of the hospital is alleviated and thus the quality and overall patient experience can be improved in the hospital.

7.3. Results of the Flat Rate Case

Now the counties are ranked by their travel burden f_i, and the optimal policy shows that any community that is further than or equal to Citrus county in distance should receive VC visits, and the rest (which include large communities such as Marion and Columbia) should come to the hospital. Marion and Columbia are two counties that are not very far from the central hospital but are close to clinics, leading to low nurse dispatch costs. Under discriminative pricing, they are offered with telehealth at a low price because of the low nurse dispatch costs. But, under non-discriminative pricing, the medical institution cannot afford to set such a low flat price for all other patients. Instead, it rather sets a higher flat price to enforce these patients to go to the hospital. Subsequently, the benefits of telehealth are deprived for the patients residing in these two communities.

Indeed, there are patients from other communities who can benefit from the flat rate policy, but the total patient surplus could be worse. The following case is an example. With a flat rate, more (52.5%) patients come to the hospital compared with the discriminative pricing case (50.5%). As a result, the surplus of medical institution, patients, and social welfare are $307K, $581K, and $888K per year, which are all smaller than those of the discriminative pricing case ($315K, $612K, and $927K). The flat-rate requirement can be less cost effective.

The social welfare maximizer lets all patients stay at home in both pricing settings, with the surplus of the medical institution, patients, and social welfare being $-3K, $616K, and $613K per year, compared to $7K, $606K, and $613K in the discriminative pricing case. In this case, the maximum feasible C_2 is chosen, but the medical institution is still worse off. The revenue gain of the medical institution is transferred to patients' aggregate surplus, and as a whole, the social welfare surplus stays the same.

8. Conclusions and Future Work

While telehealth technology and its use are not new, widespread adoption among care providers and patients beyond simple telephone correspondence has been relatively slow. The virtual services offered are typically limited to VC visits without nurse/surrogate assistance, and mainly cover chronic diseases that do not require a physical examination (mental diseases for example). However, recent policy changes during the COVID-19 pandemic have reduced barriers to telehealth access, and have promoted the use of telehealth as a way to deliver care across a broad spectrum. The involvement of nurse assistance makes it possible for VC visits to cover extended health conditions. To the best of our knowledge, very few work exists to investigate the design of VC visits with nurse assistance, and our model fills this gap to help community hospitals understand the potential of VC visits.

Our study has limitations. First, our analytic results are conducted under some strong assumptions. For instance, nurse coordination costs are assumed to be proportional to the number of patients. The cost structure and the factors influencing costs can be further investigated. Second, the heterogeneity of other patient features on top of the travel burden is not captured in the current study. For instance, the technology adoption intention of patients can change over time, and the health disparity in demographic and socio-economic features of patients may affect their willingness to use telehealth services. These personalized and possibly evolving preferences need to be considered. Currently, they are difficult to be quantified lacking related data as telehealth is in its early adoption stage.

Moreover, barriers to implementing VC visits need to be removed. How to improve patient awareness of emerging virtual care services is a challenging task and should be further investigated. Admittedly, VC visits are yet to enjoy the uptake by patients and care providers alike, both expressed security and privacy concerns of such virtual services. Another critical determinant of the success of VC visits is the payment model. Our model assumes that insurance companies would reimburse the same amount for physician services rendered by the medical institution. Currently, there is no well-established compensation model for providers who offer telehealth, nor a clear guideline of co-payment or co-insurance by patients. As Medicare recently extends its coverage to some telehealth services (U.S. Government of Medicare 2019), we expect more care plans would accept telehealth services. The current model can be extended with a refined compensation scheme. In addition, we consider a monopolistic medical institution and do not allow for patients the no-treatment option. A discussion regarding the impact of VC visits on competitive medical institutions is explored in the Appendix, and more investigation into patient loss or migration will be pursued in a future study.

There are several other future research directions. One is to extend the content of impact that telehealth is able to exert. Our model only considers the direct revenue gain from offering VC visits and does not account for the impact on the medical institution's operational cost and patient outcomes. Other clinical studies have reported provider efficiency gain (Allen et al. 1995, Stalfors et al. 2003, Olver et al. 2005), reduced appointment no-show (American Well 2019b), and reduced use of health care resources, such as hospitalizations, and primary care and emergency department visits (Polisena et al. 2009). A future study can assimilate these indirect benefits from the health economics perspective. Another direction is to explore the related operational-level optimization problems. For example, since the medical institution dispatches nurses to patient homes, the corresponding vehicle routing problem can be investigated. In addition, the nurse coordination cost is relatively easy to compute with a deterministic demand. At the strategic level, deterministic models are good "first-order" approximations (asymptotically optimal in some cases) for more sophisticated stochastic models. In our case, they provide valuable insights into how optimal pricing policies are shaped by distinct parameters of the model. At the operational level, with a stochastic demand in reality, nurses need to be dynamically assigned to different communities, and a resource allocation problem to minimize the nurse coordination cost with or without limited capacity of nurses can be considered.

References

Abrams, D. J., and Geier., M. R. (2006). A comparison of patient satisfaction with telehealth and on-site consultations a pilot study for prenatal genetic counseling. *Journal of Genetic Counseling*, 15, 199.

Allen, A., Hayes, J., Sadasivan, R., Williamson, S. K., and Wittman., C. (1995). A pilot study of the physician acceptance of tele-oncology. *Journal of Telemedicine and Telecare*, 1, 34–37.

American Well (2019a). One in five consumers would switch to a doctor that offers telehealth visits. https://www.americanwell.com/press-release/one-in-five-consumers-would-switch-to-a-doctor-that-offers-telehealth-visits/.

American Well (2019b). Telemedicine follow-up care helps reduce no-show rates. https://www.americanwell.com/telemedicine-follow-up-care-reduce-no-shows/.

AmwellHealth (2020). Telehealth and the Future of Hybrid Care. https://health.oliverwyman.com/2020/10/the-future-of-hybrid-care.html. [Online; accessed 19-May-2021].

Bavafa, H., Savin, S., and Terwiesch., C. (2018). Redesigning primary care delivery: Customized office revisit intervals and e-visits. *Available at SSRN 2363685*.

Bayram, A., Deo, S., Iravani, S., and Smilowitz., K. (2020). Managing virtual appointments in chronic care. *IISE Transactions on Healthcare Systems Engineering*, 10 (1), 1–17.

Bellot, D., Boyer, A., and Charpillet., F. (2002). Designing smart agent based telemedicine with dynamic bayesian networks: an application to kidney disease people. https://www.semanticscholar.org.

Bohnenkamp, S. K., McDonald, P., Lopez, A. M., Krupinski, E., and Blackett., A. (2004). Traditional versus telenursing outpatient management of patients with cancer with new ostomies. In *Oncology Nursing Forum*, volume 31, p. 1005. Oncology Nursing Society 125 Enterprise Dr, Pittsburgh, PA 15275 USA.

CDC (2020). Using Telehealth to Expand Access to Essential Health Services during the COVID-19 Pandemic. https://www.cdc.gov/coronavirus/2019-ncov/hcp/telehealth.html. [Online; accessed 19-May-2021].

Clegg, A., Brown, T., Engels, D., Griffin, P., and Simonds., D. (2011). Telemedicine in a rural community hospital for remote wound care consultations. *Journal of Wound Ostomy & Continence Nursing*, 38, 301–304.

Dávalos, M. E., French, M. T., Burdick, A. E., and Simmons., S. C. (2009). Economic evaluation of telemedicine: review of the literature and research guidelines for benefit–cost analysis. *Telemedicine and e-Health*, 15, 933–948.

Eriksson, L., Lindström, B., and Ekenberg., L. (2011). Patients' experiences of telerehabilitation at home after shoulder joint replacement. *Journal of Telemedicine and Telecare*, 17, 25–30.

eVisit (2020). Reimbursement medicare & telemedicine reimbursement. https://evisit.com/resources/medicare-telemedicine-reimbursement/.

Grady, B. J. (2002). A comparative cost analysis of an integrated military telemental health-care service. *Telemedicine Journal and E-health*, 8, 293–300.

Grant, L. A., Rockwood, T., and Stennes., L. (2015). Client satisfaction with telehealth services in home health care agencies. *Journal of Telemedicine and Telecare*, 21, 88–92.

Hu, P., Lu, Y., and Song., M. (2019). Joint pricing and inventory control with fixed and convex/concave variable production costs. *Production and Operations Management*, 28, 847–877.

Hunter, D. R., and Lange., K. (2000). Quantile regression via an mm algorithm. *Journal of Computational and Graphical Statistics*, 9, 60–77.

Kitamura, C., Zurawel-Balaura, L., and Wong., R. (2010). How effective is video consultation in clinical oncology? A systematic review. *Current Oncology*, 17, 17.

Lach, J. M., and Vázquez., R. M. (2004). Simulation model of the telemedicine program. In *Proceedings of the 36th Conference on Winter Simulation*, pp. 2012–2017. Winter Simulation Conference.

Laila, M., Rialle, V., Nicolas, L., Duguay, C., and Franco., A. (2008). Videophones for the delivery of home healthcare in oncology. *Studies in Health Technology and Informatics*, 136, 39.

Liu, Y., Wang, X., Gilbert, S., and Lai., G. (2018). Pricing, quality and competition at on-demand healthcare service platforms. *Available at SSRN 3253855*.

Mair, F., Whitten, P., May, C., and Doolittle., G. C. (2000). Patients' perceptions of a telemedicine specialty clinic. *Journal of Telemedicine and Telecare*, 6, 36–40.

Mallow, J. A., Petitte, T., Narsavage, G., Barnes, E., Theeke, E., Mallow, B. K., and Theeke., L. A. (2016). The use of video conferencing for persons with chronic conditions: A systematic review. *E-health telecommunication systems and networks*, 5, 39.

U.S. Government of Medicare (2019). Medicare your medicare coverage: Telehealth. https://www.medicare.gov/coverage/telehealth.

Miller, J. (2020). Why nurses are crucial to the growth of telehealth. https://healthtechmagazine.net/article/2020/06/why-nurses-are-crucial-growth-telehealth.

Olver, I., Brooksbank, M., Champion, N., and Keeley., J. (2005). The use of videophones to enhance palliative care outreach nursing in remote areas. *Progress in Palliative Care*, 13, 263–267.

Pavia, T. M. (1995). Profit maximizing cost allocation for firms using cost-based pricing. *Management Science*, 41, 1060–1072.

Polisena, J., Coyle, D., Coyle, K., and McGill., S. (2009). Home telehealth for chronic disease management: a systematic review and an analysis of economic evaluations. *International Journal of Technology Assessment in Health Care*, 25, 339–349.

Powell, R. E., Henstenburg, J. M., Cooper, G., Hollander, J. E., and Rising., K. L. (2017). Patient perceptions of telehealth primary care video visits. *The Annals of Family Medicine*, 15, 225–229.

Qian, W., Sankar, R., Song, X., Sun, X., and Clark., R. (2005). Standardization for image characteristics in telemammography using genetic and nonlinear algorithms. *Computers in Biology and Medicine*, 35, 183–196.

Rajan, B., and Agnihothri., S. (2019). Operational impact of mhealth adoption in clinical practice. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*.

Rajan, B., Tezcan, T., and Seidmann., A. (2019). Service systems with heterogeneous customers: Investigating the effect of telemedicine on chronic care. *Management Science*, 65, 1236–1267.

Saghafian, S., Hopp, W. J., Iravani, S. M., Cheng, Y., and Diermeier., D. (2018). Workload management in telemedical physician triage and other knowledge-based service systems. *Management Science*, 64 (11), 5180–5197.

Savin, S., Xu, Y., and Zhu., L. (2019). Delivering multi-specialty care via online telemedicine platforms. *Available at SSRN 3479544*.

Schmidt, K. (2020). 12 types of bill rates that affect travel nursing pay. https://blog.bluepipes.com/bill-rates-and-travel-nursing-pay/.

Sevean, P., Dampier, S., Spadoni, M., Strickland, S., and Pilatzke., S. (2009). Patients and families experiences with video telehealth in rural/remote communities in northern canada. *Journal of Clinical Nursing*, 18, 2573–2579.

Stalfors, J., Holm-Sjögren, L., Schwieler, Å., Törnqvist, H., and Westin., T. (2003). Satisfaction with telemedicine presentation at a multidisciplinary tumour meeting among patients with head and neck cancer. *Journal of Telemedicine and Telecare*, 9, 150–155.

Tarakci, H., Sharafali, M., and Ozdemir., Z. (2007). Optimal staffing policy and telemedicine. *AMCIS 2007 Proceedings*, 1, 226–232.

The Official U.S. Government Site for Medicare (2020). Medicare telehealth. https://www.medicare.gov/coverage/telehealth.

Treurnicht, M. (2009). A decision support framework for telemedicine implementation in the developing world. *Stellenbosch University, Stellenbosch*.

UF Health Cancer Center (2020). UFHCC the population we serve. https://cancer.ufl.edu/community/the-population-we-serve/.

Wood, M. (2019). Health Information Technology telemedicine to attract 7m patient users by 2018 — 12 statistics on the thriving market. https://www.beckershospitalreview.com/healthcare-information-technology/telemedicine-to-attract-7m-patient-users-by-2018-12-statistics-on-the-thriving-market.html.

Zhong, X., Hoonakker, P., Bain, P. A., Musa, A. J., and Li., J. (2018). The impact of e-visits on patient access to primary care. *Health Care Management Science*, 21 (4), 475–491.

Zhong, X., Li, J., Bain, P. A., and Musa., A. J. (2017). Electronic visits in primary care: Modeling, analysis, and scheduling policies. *IEEE Transactions on Automation Science and Engineering*, 14, 1451–1466.

Supplemental Online Materials to **Service System Design of Video Conferencing Visits with Nurse Assistance**, by Xiaojie Wang, Yongpei Guan, and Xiang Zhong.

Appendix A Summary of Model Notations

Table A.1 Summary of model notations.

Indices:	
i	$i \in I$ marks communities, $N = \|I\|$;
Parameters:	
$\mathcal{R}_1$	reward for care delivered in the hospital (\$);
$\mathcal{R}_2$	reward for care delivered at home (\$);
C_1	price charged for physician services (\$);
C_2	price charged for nursing services that medical institution cannot determine (\$);
η_d	percentage of $C_{2,i}$ paid by patients;
η_f	percentage of $C_{2,f}$ paid by patients;
D_i	demand rate in community i;
f_i	travel burden of community i patients (\$);
α	sensitivity of congestion from the patient's side (\$ per person: \$/pp);
γ	sensitivity of congestion from the medical institution's side (\$/pp);
$g_i(D)$	nurse coordination cost of community i as a function of demand D (\$);
β_i	incremental cost of the nurse coordination cost of community i (\$/pp);
θ	fixed cost of the nurse coordination cost (\$);
W_e	threshold of congestion from the medical institution's side.
Decision variables:	
$C_{2,i}$	price for VC services delivered in community i if discriminative pricing is allowed (\$);
$C_{2,f}$	flat price for VC services if discriminative pricing is not allowed (\$);
U_i	utility of patients in community i (\$);
ρ_i	proportion of community i patients going to the hospital.

Appendix B Additional Results for Two Communities

With two communities A and B, let the partitions ρ_A and ρ_B be further indexed by a superscript for different equilibria $S^{m_A m_B}$, where the scenario indicators $m_i \in \{1,2,3\}$, $i \in \{A,B\}$, inheriting the definition of that of a single community. To illustrate, when both communities go to the hospital, $m_A = m_B = 1$, and $(\rho_A^{11}, \rho_B^{11}) = (1,1)$. In the following, we assume $D_A \leq D_B$ for illustration purposes. Due to symmetry, systems with $D_A \geq D_B$ can be analyzed analogously.

Proposition A.1. *Under the nurse coordination cost structure $g_i(D) = \beta_i D + \theta I_{D>0}$, with an equal marginal utility change ($\varphi_{\mathbf{A}} = \varphi_{\mathbf{B}} = \varphi$), the optimal patient diversion policies for two communities are as follows:*

1. *Both staying at home (S^{22}), if* $\max\{\alpha(D_A + D_B), 2\theta/(D_A + D_B)\} \leq \varphi$.

2. Both being partially diverted (S^{33}), with proportions $(\rho_A^{33}, \rho_B^{33})$ going to the hospital, where $\rho_B^{33} D_B + \rho_A^{33} D_A = (D_A + D_B)/2 - \varphi/(2\alpha)$, if $\max\{2\sqrt{2\alpha\theta} - \alpha(D_A + D_B), \alpha(D_B - D_A)\} \leq \varphi \leq \alpha(D_A + D_B)$. The special cases S^{23} and S^{32} belong to this region.

3. $(1, \rho_B^{13})$ or $(\rho_A^{31}, 1)$, where $\rho_B^{13} = (D_B - D_A)/(2D_B) - \varphi/(2\alpha D_B)$, and $\rho_A^{31} = (D_A - D_B)/(2D_A) - \varphi/(2\alpha D_A)$, i.e., community A (respectively, B) patients go to the hospital and a mix strategy is adopted for community B (respectively, A), if $-\alpha(D_A + D_B) + 2\sqrt{\alpha\theta} \leq \varphi \leq \alpha(D_B - D_A)$, or $-\alpha(D_A + D_B) + 2\sqrt{\alpha\theta} \leq \varphi \leq \alpha(D_A - D_B)$, correspondingly. Note that this region is feasible only when $\theta < \alpha D_B^2$.

4. Both going to the hospital (S^{11}) for all other parameter regions.

Proposition A.2 summarizes the optimal patient diversions in different parameter regions.

PROPOSITION A.2. *Under the nurse coordination cost structure $g_i(D) = \beta_i D + \theta I_{D>0}$, for two communities with unequal φ's, the following scenarios can unfold:*

1. $\theta = 0$. The optimal equilibria are summarized in Table A.2 and the partitions are presented in Figure 3(a).

2. $0 \leq \theta < \alpha D_A^2$. Similar optimal equilibria exist but in different parameter regions, as shown in Figure 3(b).

3. $\alpha D_A^2 \leq \theta < \alpha D_B^2$. Compared to Figures 3(a) and (b), equilibria S^{32} and S^{31} no longer exist, see Figure 3(c).

4. $\alpha D_B^2 \leq \theta$. No mixed strategy exists, see Figure 3(d).

Table A.2 Optimal patient diversions and the corresponding parameter regions when $\varphi_A \neq \varphi_B$.

Scenario	Patient diversion		Parameter region ($\theta = 0$)
S^{11}	$(1,1)$		$\zeta_A \leq -\frac{D_A+D_B}{2D_A}$ and $\zeta_B \leq -\frac{D_A+D_B}{2D_B}$
S^{12}	$(1,0)$		$\zeta_A \leq \frac{D_B-D_A}{2D_A}$ and $\zeta_B \geq \frac{D_B-D_A}{2D_B}$
S^{13}	$(1,\rho_B^{13})$	$\rho_B^{13} = \frac{D_B-D_A}{2D_B} - \zeta_B$	$\frac{-D_A-D_B}{2D_B} \leq \zeta_B \leq \frac{D_B-D_A}{2D_B}$ and $\zeta_B D_B \geq \zeta_A D_A$
S^{21}	$(0,1)$		$\zeta_A \geq \frac{D_A-D_B}{2D_A}$ and $\zeta_B \leq -\frac{D_A+D_B}{2D_B}$
S^{22}	$(0,0)$		$\zeta_A \geq \frac{D_A+D_B}{2D_A}$ and $\zeta_B \geq \frac{D_A+D_B}{2D_B}$
S^{23}	$(0,\rho_B^{23})$	$\rho_B^{23} = \frac{D_A+D_B}{2D_B} - \zeta_B$	$\frac{D_A-D_B}{2D_B} \leq \zeta_B \leq \frac{D_A+D_B}{2D_B}$ and $\zeta_B D_B \leq \zeta_A D_A$
S^{31}	$(\rho_A^{31},1)$	$\rho_A^{31} = \frac{D_A-D_B}{2D_A} - \zeta_A$	$\frac{-D_A-D_B}{2D_A} \leq \zeta_A \leq \frac{D_A-D_B}{2D_A}$ and $\zeta_B D_B \leq \zeta_A D_A$
S^{32}	$(\rho_A^{32},0)$	$\rho_A^{32} = \frac{D_A+D_B}{2D_A} - \zeta_A$	$\frac{D_B-D_A}{2D_A} \leq \zeta_A \leq \frac{D_A+D_B}{2D_A}$ and $\zeta_B D_B \geq \zeta_A D_A$

Appendix C Discussion on Competitors

When competition is induced in the market, it is natural that patients would benefit from having more choices. Here we shed light on the impact of competition on the medical institution. Consider two adjacent service regions, one is covered by medical institution M_A and the other M_B. Prior to virtual services, there is no patient migration due to a high travel cost. For instance, when the two service regions are of the same size, and thus the same level of congestion in the hospital, patients will always choose the closer one to visit. If medical institution M_B is the only one that offers VC visits, it can "snatch" patients from medical institution M_A, which is a trivial case. We further consider both medical institutions start offering virtual care services.

Under this setting, region A patients actually face four choices, (1) go to the hospital A, (2) go to the hospital B, (3) VC visits provided by M_A, and (4) VC visits provided by M_B. Option 2 remains feasible for region A patients living in the borderline area of two service regions, when the congestion of hospital B is reduced significantly due to M_B's virtual service. Then, whoever can offer VC visits at a lower price becomes the winner of this competition. If M_B offers a price $C_1^{M_B} + \eta_d C_{2,i}^{M_B} + C_2^{M_B}$ such that M_A cannot match, i.e., medical institution M_A will lose money to serve patients priced similarly to M_B, then, some patients in region A will choose services from M_B until the capacity cap of M_B is reached. Because of the community-dependent nurse coordination cost, some communities in region B can also benefit from services provided by M_A, where a bi-directional patient migration could happen. Even without losing patients, the medication institutions still have to lower prices because they can no longer leverage "artificial congestion" by requesting a portion of patients to come to hospital as before. The competitor M_B can play as the tie-breaker to make going to the hospital A strictly worse than receiving VC visits provided by M_B. When being the monopoly, M_A can freely set VC visit prices to benchmark in-person services, and the prices make it profitable to divert patients to stay at home. However, for the competitor M_B, aimed at "snatching" patients, it does not hold the same benchmark as serving patients in A hospital, and as far as priced higher than its operating cost and under its capacity limits, it is willing to set prices as low as possible to attract patients. Therefore, if there exists a migration from community i in service region A to B, the harm made to M_A could be greatly larger than the benefit gained by M_B.

It is possible that the two medical institutions as a whole is worse off compared to not having virtual services. In particular, define π^{M_A} and π^{M_B} as the basic costs of offering care services. They are not included in the monopoly case because they are kept the same, regardless of the extra cost of nurse coordination when in a virtual way, but they come into play here with competitions. Let $\beta_i^{M_A}$ and $\beta_i^{M_B}$ be the marginal nurse coordination costs. Then, due to the loss of D_i patients to hospital B, the difference in profit $\Delta P^A = (\pi^{M_A} - C_1^{M_A} + \gamma^{M_A})D_i < 0$, and that of M_B, $\Delta P^B = (C_1^{M_B} + \eta_d C_{2,i}^{M_B} + C_2^{M_B} - \pi^{M_B} - \beta_i^{M_B})D_i \geq 0$. When M_B is willing to set $C_1^{M_B} + \eta_d C_{2,i}^{M_B} + C_2^{M_B} = \pi^{M_A} + \beta_i^{M_A}$ to attract M_A patients, then, the total change $\Delta P^A + \Delta P^B = (\beta_i^{M_A} - \beta_i^{M_B} + \pi^{M_A} - C_1^{M_A} + \gamma^{M_A})D_i$, given $\pi^{M_A} = \pi^{M_B}$, $C_1^{M_A} = C_1^{M_B}$ and $\gamma^{M_A} = \gamma^{M_B}$ to mimic a symmetric case. The change in profit as a whole depends on the difference of the nurse coordination costs between the two medical institutions. If the gain from the saving of nurse coordination costs is not large, but the original price $C_1^{M_A}$ is much higher than cost π^{M_A}, then VC visits can be detrimental.

The analysis above suggests that VC visits could break the geographic "quarantine" that stabilizes the market, and stimulate competition among care providers. It further stresses that the competition might induce unintended consequences to medical institutions. On patients, based on the utility definition, they are no worse off, however, when patients switch between multiple providers to reduce cost, the continuity of care can be affected, which is not good for patient health management in the long run. To avoid patient loss, medical institutions need to provide high-quality care, both in-person and virtually, to improve patient retention and maintain competitiveness in the market.

Appendix D Discussion on Payer Reimbursement

As an extension of the basic payment structure for the discriminative pricing model, we further consider that the insurer will reimburse $\hat{C}_1$ for the physician's portion, and $\delta C_{2,i}$ for the nursing service, where $C_{2,i}$ is the billed price. An analysis of the non-discriminative case can be carried out similarly. As such, the medical institution will collect $C_1^M = C_1 + \hat{C}_1$ for office visits and $C_{2,i}^M = C_1 + \hat{C}_1 + (\delta + \eta_d)C_{2,i} + C_2$ for VC visits. When the service capacity is fixed, the term $C_1 + \hat{C}_1$ is immaterial, and the objective function becomes:

$$R^\delta = \max_{C_{2,i}} \sum_i D_i(1-\rho_i)((\eta_d + \delta)C_{2,i} + C_2) - \gamma \left[\sum_i \rho_i D_i - W_e\right]^+ - \sum_i g_i((1-\rho_i)D_i).$$

The introduction of payer reimbursement will not affect the IC constraints of patients. Under this circumstance, the system-level marginal gain φ_i should be modified as $\varphi_i^\delta = -\Delta\mathcal{R} + f_i - \delta C_2/(\eta_d + \delta) - \eta_d\beta_i/(\eta_d+\delta) + \eta_d\gamma/(\eta_d+\delta)$, and $R^\delta = \max_{C_{2,i}} \sum_i D_i(1-\rho_i)(\varphi_i^\delta + \alpha W)(\eta_d+\delta)/\eta_d$. The equilibrium patient partition follows the same formulation except that φ_i is replaced by φ_i^δ. In other words, a threshold-type of equilibrium strategy of patients is invariant to the amount paid by the insurer to the medical institution.

However, the relationship between the co-insurance factor η_d and the reimbursement factor δ does affect the medical institution's profitability. Given a fixed amount collected by the medical institution per patient, e.g., $\eta_d + \delta = 1$, the larger δ values, the more revenue the medical institution will gain. In our numerical study, for the linear nurse cost function case, compared to the baseline scenario where $\eta_d = 1$, $\delta = 0$, and $R = \$157.5K$, a scenario of $\eta_d = 0.5$ and $\delta = 0.5$ leads to $R = \$316K$, and $\eta_d = 0.3$ and $\delta = 0.7$ leads to $R = \$528.5K$. This suggests that shifting the burden from patients to the payer (e.g., the government or insurance companies) can benefit both patients and the medical institution. It heightens the importance of the insurance coverage of telehealth related services so as to promote the adoption of telehealth by patients and care providers, and ultimately, improving care access and patient outcomes.

Appendix E Proofs

Unless otherwise notified, we consider $\eta_d \neq 0$ and C_2 is a constant in Model 1 for the basic proofs.

Lemma A.1. *Given the patient diversion ρ_i's, the equilibrium pricing for the revenue maximization problem is $\eta_d C_{2,i} = -C_2 - \Delta\mathcal{R} + f_i + \alpha W$, where $W = \sum_i \rho_i D_i$.*

Proof. This directly follows patients' IC constraints. ■

Proposition A.3. *Consider the optimization problem (Model 2) where the medical institution determines the price $C_{2,i}$ for nursing services provided in community i (patient home), and dictates a proportion $\rho_i \in [0,1]$ of patients to receive in-person visits versus VC visits (1-ρ_i):*

$$\max_{\rho_i, C_{2,i}} \sum_i D_i(1-\rho_i)(\eta_d C_{2,i} + C_2) - \gamma \left[\sum_i \rho_i D_i - W_e\right]^+ - \sum_i g_i((1-\rho_i)D_i).$$

The solutions $C_{2,i}$'s to Model 2 are also the optimal solutions to Model 1.

Proof. Here we aim to prove that the optimal solutions of Model 1 and Model 2 are identical, and the medical institution can use discriminatory pricing alone to control patients' diversion decisions.

First, the objective of Model 2 is greater than or equal to that of Model 1, because the hospital can control patients' decisions and the decision space is larger. To prove that the equivalency can be achieved, we introduce a small amount $\epsilon > 0$.

Denote the the optimal solution of Model 2 as ρ_i^2's, and the corresponding optimal congestion as $W^2 = \sum_i \rho_i^2 D_i$, the hospital can set the at-home price for each community i as:

(1) $\eta_d C_{2,i}^* + C_2 = -\Delta\mathcal{R} + f_i + \alpha W^2 + \epsilon$, for community i such that $\rho_i^2 = 1$,

(2) $\eta_d C_{2,i}^* + C_2 = -\Delta\mathcal{R} + f_i + \alpha W^2$, for community i such that $\rho_i^2 \in (0,1)$,

(3) $\eta_d C_{2,i}^* + C_2 = -\Delta\mathcal{R} + f_i + \alpha W^2 - \epsilon$, for community i such that $\rho_i^2 = 0$,

where $\Delta\mathcal{R} = \mathcal{R}_1 - \mathcal{R}_2$.

First, consider $\eta_d \neq 0$. Note that $\{C_{2,i}^*\}$'s are also feasible solutions to Model 1. In this way, according to incentive compatibility, in equilibrium, patients in case (1) naturally choose to go to the hospital (i.e., $\rho_i^1 = \rho_i^2 = 1$) and those in case (3) naturally choose to stay at home (i.e., $\rho_i^1 = \rho_i^2 = 0$). Eventually, the hospital can use $\{C_{2,i}^*\}$'s to determine the congestion (i.e., the total number of patients coming to the hospital W^1), and in equilibrium, $W^1 = W^2$. Since the hospital is optimal to set at most one $\rho_i^2 \in (0,1)$ based on Model 2 (Theorem 2), then, there will be at most one community in case (2). Denote it as community k, then, $W^1 = W^2$ will induce $\rho_k^1 = \rho_k^2$.

Next, consider $\eta_d = 0$ and $\eta_f \neq 0$.Following the same logic with $\eta_d \neq 0$, we have Model 1 as

$$\max_{C_{2,f}} (\eta_f C_{2,f} + C_2)\left(\sum_{j:f_j > f_i} D_j + (1-\rho_i)D_i\right) - \sum_{j:f_j > f_i} g_j(D_j) - g_i((1-\rho_i)D_i) - \gamma\left[\sum_{j:f_j < f_i} D_j + \rho_i D_i - W_e\right]^+. \tag{A.1}$$

Model 2 is

$$\max_{C_{2,f},\rho_i} (\eta_f C_{2,f} + C_2)\left(\sum_{j:f_j > f_i} D_j + (1-\rho_i)D_i\right) - \sum_{j:f_j > f_i} g_j(D_j) - g_i((1-\rho_i)D_i) - \gamma\left[\sum_{j:f_j < f_i} D_j + \rho_i D_i - W_e\right]^+. \tag{A.2}$$

Denote the optimal solution of Model 2 as ρ_i^2's, and the corresponding optimal congestion as $W^2 = \sum_i \rho_i^2 D_i$, the hospital can set the at-home price for each community i as: $\eta_f C_{2,f} + C_2 = -\Delta\mathcal{R} + f_i + \alpha W^2$, for community i such that $\rho_i^2 \in (0,1)$, where $\Delta\mathcal{R} = \mathcal{R}_1 - \mathcal{R}_2$.

Then, for $f_j > f_i$, we have $\eta_f C_{2,f} + C_2 < -\Delta\mathcal{R} + f_j + \alpha W^2$, so patients in community j would naturally stay at home; for $f_j < f_i$, we have $\eta_f C_{2,f} + C_2 > -\Delta\mathcal{R} + f_j + \alpha W^2$, so patients in community j would automatically go to hospital. Therefore, Model 1 and Model 2 are equivalent. ∎

All of the results below are derived from Model 2 instead of Model 1 for its simplicity.

Proof of Proposition 1

Proof. For one community, three scenarios can unfold: (1) all patients go to the hospital, (2) all patients stay at home, and (3) part of them stay at home. The corresponding objective value is denoted as R^m, $m \in \{1,2,3\}$.

For $\rho \in [0,1]$, based on Lemma A.1, we have the medical institution's revenue as:

$$R = D(1-\rho)(\Delta\mathcal{R} + f + \alpha\rho D) - \gamma\rho D - \beta(1-\rho)D - \theta I_{(1-\rho)D>0}.$$

When $\rho \in (0,1)$, the first-order condition (FOC) $\frac{\partial R}{\partial \rho} = -(\varphi + \alpha\rho D)D + \alpha(1-\rho)D^2 = 0$ implies $\rho = \frac{1}{2} - \frac{\varphi}{2\alpha D} := \frac{1}{2} - \zeta$, where $\zeta = \frac{\varphi}{2\alpha D}$. Then, $R^3 - R^1 = D(1-\rho)(\varphi + \alpha\rho D) - \theta = \alpha D^2(\frac{1}{2} + \zeta)^2 - \theta$. The partition line between R^3 and R^1 is $\zeta = -\frac{1}{2} + \frac{\sqrt{\theta}}{\sqrt{\alpha}D}$. Furthermore, $R^3 - R^2 = \alpha D^2(\frac{1}{2} - \zeta)\rho$. The partition line between R^3 and R^2 is invariant with θ, which is $\zeta = \frac{1}{2}$. Thus, when $\theta \le \alpha D^2$, R^3 is optimal when $-\frac{1}{2} + \frac{\sqrt{\theta}}{\sqrt{\alpha}D} \le \zeta \le \frac{1}{2}$, R^1 is optimal when $\zeta < -\frac{1}{2} + \frac{\sqrt{\theta}}{\sqrt{\alpha}D}$, and R^2 is optimal when $\frac{1}{2} < \zeta$. When $\theta \ge \alpha D^2$, R^3 is dominated by R_1. The partition line between the two scenarios with $R^2 - R^1 = \varphi D - \theta = 0$ is $\zeta = \frac{\theta}{2\alpha D^2}$. R^1 is optimal when $\zeta < \frac{\theta}{2\alpha D^2}$ and R^2 is optimal when $\zeta > \frac{\theta}{2\alpha D^2}$.
∎

Proof of Proposition A.1

Proof. For illustration purposes, we assume $D_B \ge D_A$. Due to the symmetry of D_A and D_B, $D_B < D_A$ can be analyzed analogously. A total of nine scenarios can occur, denoted as S^{mn}, and the corresponding revenue is denoted as R^{mn}, $m,n \in \{1,2,3\}$, similar to that defined in the proof of Proposition 1. Let ρ_i's be further indexed by a superscript mn, $m,n \in \{1,2,3\}$ to refer to the diversion under the scenario S^{mn}. To illustrate, by definition, when both communities go to the hospital, $(\rho_A^{11}, \rho_B^{11}) = (1,1)$. We first introduce the revenue obtained when both communities have patients staying at home, i.e., scenarios S^{33}, S^{32}, and S^{23}.

$$R = (1-\rho_A)D_A(-\Delta\mathcal{R} + f_A + \alpha(\rho_A D_A + \rho_B D_B)) \tag{A.3}$$

$$+(1-\rho_B)D_B(-\Delta\mathcal{R}+f_B+\alpha(\rho_A D_A+\rho_B D_B))$$
$$-\gamma(\rho_A D_A+\rho_B D_B)-\beta_A(1-\rho_A)D_A-\beta_B(1-\rho_B)D_B-2\theta.$$

Introduce the marginal utility change as $\varphi_i = f_i - \beta_i + \gamma - \Delta\mathcal{R},\ i \in \{A, B\}$. The Lagrange dual is $L = R + \lambda_A\rho_A + \mu_A(1-\rho_A) + \lambda_B\rho_B + \mu_B(1-\rho_B)$, where $\lambda_A, \lambda_B, \mu_A, \mu_B \geq 0$, and the FOC is as follows:

$$\frac{\partial L}{\partial \rho_A} = -(\varphi_A + \alpha\rho_A D_A + \alpha\rho_B D_B)D_A + \alpha(1-\rho_A)D_A^2 + \alpha(1-\rho_B)D_B D_A + \lambda_A - \mu_A = 0,$$
$$\frac{\partial L}{\partial \rho_B} = -(\varphi_B + \alpha\rho_A D_A + \alpha\rho_B D_B)D_B + \alpha(1-\rho_B)D_B^2 + \alpha(1-\rho_A)D_A D_B + \lambda_B - \mu_B = 0.$$

When $\rho_A, \rho_B \in (0,1)$, the FOC is feasible if and only if $\varphi_A = \varphi_B = \varphi$, and

$$\rho_B D_B + \rho_A D_A = -\frac{\varphi}{2\alpha} + \frac{D_A + D_B}{2}. \tag{A.4}$$

In this case, there can be multiple solutions of ρ_A and ρ_B, as long as equation (A.4) and $\alpha(D_A + D_B) \geq \varphi$ hold. Some special cases include $(\rho_A^{32}, 0)$ with $\rho_A^{32} = \frac{D_A+D_B}{2D_A} - \frac{\varphi}{2\alpha D_A}$, and $(0, \rho_B^{23})$ with $\rho_B^{23} = \frac{D_A+D_B}{2D_B} - \frac{\varphi}{2\alpha D_B}$.

In addition, the revenue when there is only one community having partial patients staying at home can be derived with reducing 2θ to θ in formula (A.3), and the FOC stays the same. As long as equation (A.4) and $\alpha(D_B - D_A) \geq \varphi$ hold, $(1, \rho_B^{13})$ with $\rho_B^{13} = \frac{D_B-D_A}{2D_B} - \frac{\varphi}{2\alpha D_B}$ is the solution, and when $\alpha(D_A - D_B) \geq \varphi$, the same is true for $(\rho_A^{31}, 1)$ with $\rho_A^{31} = \frac{D_A-D_B}{2D_A} - \frac{\varphi}{2\alpha D_A}$.

For the trivial cases, we present the revenues in Table A.3. Next, we compare and decide which scenario will lead to the largest revenue with a given parameter region.

Table A.3 Revenues under different scenarios when $\varphi_A = \varphi_B$.

Patient diversion	Revenue
(1,1)	$R^{11} = (D_A + D_B)(-\gamma)$
(0,1)	$R^{12} = R^{11} + D_A(\varphi + \alpha D_B) - \theta$
(1,0)	$R^{21} = R^{11} + D_B(\varphi + \alpha D_A) - \theta$
(0,0)	$R^{22} = R^{11} + \varphi(D_A + D_B) - 2\theta$
$(\rho_A^{33}, \rho_B^{33})$	$R^{33} = R^{11} + \frac{1}{\alpha}(\frac{\varphi}{2} + \frac{\alpha(D_A+D_B)}{2})^2 - 2\theta$

We start with $\theta = 0$ to understand the solution structure. In this no set-up cost case, in the region of $|\varphi| < \alpha(D_A + D_B)$, as long as ρ_A and ρ_B satisfy equation (A.4), the revenue stays the same, including scenarios $S^{13}, S^{31}, S^{23}, S^{32}$ and S^{33}. We choose the mixed strategy $(\rho_A^{33}, \rho_B^{33})$ and we claim that $R^{33} \geq R^{22} \vee R^{21} \vee R^{12} \vee R^{11}$. This follows the pairwise comparisons: $R^{33} - R^{22} = \frac{1}{\alpha}(\frac{\varphi}{2} - \frac{\alpha(D_A+D_B)}{2})^2 \geq 0$; $R^{33} - R^{21} = \frac{1}{\alpha}(\frac{\varphi}{2} + \frac{\alpha(D_A-D_B)}{2})^2 \geq 0$; $R^{33} - R^{12} = \frac{1}{\alpha}(\frac{\varphi}{2} + \frac{\alpha(-D_A+D_B)}{2})^2 \geq 0$; and $R^{33} - R^{11} = \frac{1}{\alpha}(\frac{\varphi}{2} + \frac{\alpha(D_A+D_B)}{2})^2 \geq 0$.

In the region $\varphi \geq \alpha(D_A + D_B)$, note that $(\rho_A^{33}, \rho_B^{33})$ is infeasible, we then claim that R^{22} is optimal in the region: $\varphi > 0$ leads to $R^{22} > R^{11}$, $\varphi > \alpha D_B$ leads to $R^{22} > R^{21}$, and $\varphi > \alpha D_A$ leads to $R^{22} > R^{12}$.

In the region $\varphi \leq -\alpha(D_A + D_B)$, again, $(\rho_A^{33}, \rho_B^{33})$ is infeasible, and we prove that R^{11} is optimal in the region: $\varphi < 0 \Rightarrow R^{11} > R^{22}$, $\varphi < -\alpha D_A \Rightarrow R^{11} > R^{21}$, and $\varphi < -\alpha D_B \Rightarrow R^{11} > R^{12}$.

For any non-negative θ, the same derivation applies and the detailed algebraic operations are omitted. However, due to the discontinuity when changing from option 1 — go to the hospital, to option 2 or 3 — dispatch nurses, the set-up cost comes into play, and the optimal region for S^{31} and S^{13} cannot be combined with that of S^{33}. The change is demonstrated in Figure 2 in the main context. ∎

Proof of Proposition A.2

Proof. For two communities with $\varphi_A \neq \varphi_B$, we also start with the scenario $(\rho_A^{33}, \rho_B^{33})$ and the FOC is the same as that of equal φ's. Since $\varphi_A \neq \varphi_B$, S^{33} is not feasible. The optimal patient diversions corresponding to the remaining eight scenarios are presented in Table A.4.

Table A.4 Optimal patient diversions and the corresponding parameter regions when $\varphi_A \neq \varphi_B$.

Scenario	Patient diversion		Parameter region $(\theta = 0)$
S^{11}	$(1,1)$		$\zeta_A \leq -\frac{D_A+D_B}{2D_A}$ and $\zeta_B \leq -\frac{D_A+D_B}{2D_B}$
S^{12}	$(1,0)$		$\zeta_A \leq \frac{D_B-D_A}{2D_A}$ and $\zeta_B \geq \frac{D_B-D_A}{2D_B}$
S^{13}	$(1,\rho_B^{13})$	$\rho_B^{13} = \frac{D_B-D_A}{2D_B} - \zeta_B$	$\frac{-D_A-D_B}{2D_B} \leq \zeta_B \leq \frac{D_B-D_A}{2D_B}$ and $\zeta_B D_B \geq \zeta_A D_A$
S^{21}	$(0,1)$		$\zeta_A \geq \frac{D_A-D_B}{2D_A}$ and $\zeta_B \leq -\frac{D_A+D_B}{2D_B}$
S^{22}	$(0,0)$		$\zeta_A \geq \frac{D_A+D_B}{2D_A}$ and $\zeta_B \geq \frac{D_A+D_B}{2D_B}$
S^{23}	$(0,\rho_B^{23})$	$\rho_B^{23} = \frac{D_A+D_B}{2D_B} - \zeta_B$	$\frac{D_A-D_B}{2D_B} \leq \zeta_B \leq \frac{D_A+D_B}{2D_B}$ and $\zeta_B D_B \leq \zeta_A D_A$
S^{31}	$(\rho_A^{31},1)$	$\rho_A^{31} = \frac{D_A-D_B}{2D_A} - \zeta_A$	$\frac{-D_A-D_B}{2D_A} \leq \zeta_A \leq \frac{D_A-D_B}{2D_A}$ and $\zeta_B D_B \leq \zeta_A D_A$
S^{32}	$(\rho_A^{32},0)$	$\rho_A^{32} = \frac{D_A+D_B}{2D_A} - \zeta_A$	$\frac{D_B-D_A}{2D_A} \leq \zeta_A \leq \frac{D_A+D_B}{2D_A}$ and $\zeta_B D_B \geq \zeta_A D_A$

When $\varphi_i > \alpha(D_A + D_B)$, $\rho_i = 0$, i.e., the choice is to stay at home, whereas when $\varphi_i < -\alpha(D_A + D_B)$, $\rho_i = 1$, i.e., go to the hospital, regardless of the optimal solution adopted by other communities.

Therefore, in what follows, we consider the non-trivial cases. We assume $D_B \geq D_A$. Due to the symmetry of D_A and D_B, $D_B < D_A$ can be analyzed analogously.

No Set-up Cost.

The revenues under different patient diversions are presented in Table A.5. We compare and decide which scenario will lead to the largest revenue given the parameter region.

Table A.5 Revenues under different patient diversions when $\varphi_A \neq \varphi_B$.

Patient diversion	Revenue
(1,0)	$R^{12} = R^{11} + D_B(\varphi_B + \alpha D_A) - \theta$
(1,ρ)	$R^{13} = R^{11} + \frac{1}{\alpha}(\frac{\varphi_B}{2} + \frac{\alpha(D_A+D_B)}{2})^2 - \theta$
(0,1)	$R^{21} = R^{11} + D_A(\varphi_A + \alpha D_B) - \theta$
(0,0)	$R^{22} = R^{11} + D_A\varphi_A + D_B\varphi_B - 2\theta$
(0,ρ)	$R^{23} = R^{11} + (\frac{D_B-D_A}{2} + \frac{\varphi_B}{2\alpha})(\frac{\alpha(D_A+D_B)}{2} + \frac{\varphi_B}{2})$
	$+D_A(\varphi_A + \frac{\alpha(D_A+D_B)}{2} - \frac{\varphi_B}{2}) - 2\theta$
(ρ,1)	$R^{31} = R^{11} + \frac{1}{\alpha}(\frac{\varphi_A}{2} + \frac{\alpha(D_A+D_B)}{2})^2 - \theta$
(ρ,0)	$R^{32} = R^{11} + (\frac{D_A-D_B}{2} + \frac{\varphi_A}{2\alpha})(\frac{\alpha(D_A+D_B)}{2} + \frac{\varphi_A}{2})$
	$+D_B(\varphi_B + \frac{\alpha(D_A+D_B)}{2} - \frac{\varphi_A}{2}) - 2\theta$

- Region 1: $\zeta_A \leq -\frac{D_A+D_B}{2D_A}$ and $\zeta_B \leq -\frac{D_A+D_B}{2D_B}$, i.e., $\varphi_A \leq -\alpha(D_A+D_B)$ and $\varphi_B \leq -\alpha(D_A+D_B)$.

  In this region, S^{32}, S^{23}, S^{13}, and S^{31} are infeasible. Moreover,

  - $\varphi_A, \varphi_B \leq 0 \Rightarrow R^{11} \geq R^{22}$;
  - $\varphi_B \leq -\alpha D_A \Rightarrow R^{11} \geq R^{12}$;
  - $\varphi_A \leq -\alpha D_B \Rightarrow R^{11} \geq R^{21}$.

Therefore, R^{11}, i.e., the patient diversion (1,1) is optimal.

- Region 2: $\zeta_A \geq \frac{D_A+D_B}{2D_A}$ and $\zeta_B \geq \frac{D_A+D_B}{2D_B}$, i.e., $\varphi_A \geq \alpha(D_A+D_B)$ and $\varphi_B \geq (D_A+D_B)$.

  In this region, S^{32}, S^{23}, S^{13}, and S^{31} are infeasible. Moreover,

  - $\varphi_A, \varphi_B \geq 0 \Rightarrow R^{22} \geq R^{11}$;
  - $\varphi_A \geq \alpha D_B \Rightarrow R^{22} \geq R^{12}$;
  - $\varphi_B \geq \alpha D_A \Rightarrow R^{22} \geq R^{21}$.

Therefore, R^{22}, i.e., the patient diversion (0,0) is optimal.

- Region 3: $\zeta_A \leq \frac{D_B-D_A}{2D_A}$ and $\zeta_B \geq \frac{D_B-D_A}{2D_B}$, i.e., $\varphi_A \leq \alpha(D_B-D_A)$ and $\varphi_B \geq \alpha(D_B-D_A)$.

  In this region, S^{32} and S^{13} are infeasible. Moreover,

  - $\alpha D_B \geq \varphi_A \Rightarrow R^{12} \geq R^{22}$.
  - $\varphi_B + \alpha D_A \geq 0 \Rightarrow R^{12} \geq R^{11}$.

- $D_B\varphi_B \geq D_A\varphi_A \Rightarrow R^{12} \geq R^{21}$.

- In this region, we have $\varphi_B \geq \alpha(D_B - D_A)$ and $-\alpha(D_A + D_B) \leq \varphi_A \leq \alpha(D_A - D_B)$. As $R^{12} - R^{31}$ is a quadratic function of φ_A with a negative second-order coefficient, the minimum is obtained at the two endpoints of the region of interest. For one endpoint $\varphi_A = -\alpha(D_A + D_B)$, $R^{12} - R^{31} \geq \alpha D_B^2 \geq 0$, and for the other endpoint $\varphi_A = \alpha(D_A - D_B)$, $R^{12} - R^{31} \geq \alpha(D_B^2 - D_A^2) \geq 0$. Both are non-negative, and thus $R^{12} - R^{31} \geq 0$ in the region.

- We also have $\alpha(D_B - D_A) \leq \varphi_B \leq \alpha(D_A + D_B)$ and $\varphi_A \leq \alpha(D_B - D_A)$. As $R^{12} - R^{23}$ is a quadratic function of φ_B with a negative second-order coefficient, the minimum is obtained at the two endpoints. For one endpoint $\varphi_B = \alpha(D_B - D_A)$, $R^{12} - R^{23} \geq 2\alpha D_A D_B \geq 0$, and for the other endpoint $\varphi_B = \alpha(D_A + D_B)$, $R^{12} - R^{23} \geq \alpha D_A(2D_B - D_A) \geq 0$. Both are non-negative, and thus $R^{12} - R^{23} \geq 0$ in the region.

Therefore, R^{12}, i.e., the patient diversion (1,0) is optimal.

- Region 4: $\zeta_A \geq \frac{D_A - D_B}{2D_A}$ and $\zeta_B \leq -\frac{D_A + D_B}{2D_B}$, i.e., $\varphi_B \leq \alpha(D_A - D_B)$ and $\varphi_A \geq \alpha(D_A - D_B)$.

In this region, S^{23} and S^{31} are infeasible. Moreover,

- $R^{21} \geq R^{22}$: $\alpha D_A \geq \varphi_B \Rightarrow R^{21} \geq R^{22}$.

- $R^{21} \geq R^{11}$: $\varphi_A + \alpha D_B \geq 0 \Rightarrow R^{21} \geq R^{11}$.

- $R^{21} \geq R^{12}$: $R^{21} \geq \alpha D_A^2 \geq \alpha(2D_A - D_B)D_B \geq R^{12}$.

- We have $\varphi_A \geq \alpha(D_A - D_B)$ and $-\alpha(D_A + D_B) \leq \varphi_B \leq \alpha(D_A - D_B)$. As $R^{21} - R^{13}$ is a quadratic function of φ_B with a negative second-order coefficient, the minimum is obtained at the two endpoints of the region under consideration. For one endpoint $\varphi_B = -\alpha(D_A + D_B)$, $R^{21} - R^{13} \geq \alpha D_A^2 \geq 0$, and for the other endpoint $\varphi_B = \alpha(D_A - D_B)$, $R^{21} - R^{13} \geq \alpha(D_A^2 - D_A^2) = 0$. Thus, the minimum is non-negative, so $R^{21} - R^{13} \geq 0$ in the region.

- We have $\alpha(D_B - D_A) \leq \varphi_A \leq \alpha(D_A + D_B)$ and $\varphi_B \leq \alpha(D_B - D_A)$. As $R^{21} - R^{32}$ is a quadratic function of φ_A with a negative second-order coefficient, the minimum is obtained at the two endpoints of this region. For the one endpoint $\varphi_B = \alpha(D_B - D_A)$, $R^{21} - R^{32} \geq -\alpha D_A^2 + 2\alpha D_A D_B + \alpha D_B^2 \geq 0$ (because $D_B \geq D_A$). For the other endpoint $\varphi_B = \alpha(D_A + D_B)$, $R^{21} - R^{32} \geq \alpha D_B^2 \geq 0$. Thus, the minimum is non-negative, so $R^{21} - R^{32} \geq 0$ in the region.

Therefore, R^{21}, i.e., the patient diversion (0,1) is optimal.

- Region 5: $\alpha(D_B - D_A) \leq \varphi_A \leq \alpha(D_A + D_B)$ and $\varphi_B \geq \varphi_A$.

First, we notice that S^{13} and S^{31} are infeasible in this region. Moreover,

- $R^{32} > R^{22}$: $R^{32} - R^{22} = D_A(-\alpha\rho_A^2 D_A + \rho_A(-\varphi_A + \alpha D_A + \alpha D_B))$, which obtains its maximum $(-\varphi_A + \alpha D_A + \alpha D_B)^2/(4\alpha D_A)$ at $(-\varphi_A + \alpha D_A + \alpha D_B)/(2\alpha D_A)$. Since $\varphi_A \leq \alpha(D_A + D_B)$, we conclude that $\rho_A^* > 0$ and the maximum is strictly positive, i.e., $R^{32} > R^{22}$.

- $R^{32} > R^{12}$: $R^{32} - R^{12} = D_A(-\alpha\rho_A^2 D_A + \rho_A(-\varphi_A + \alpha D_A + \alpha D_B) + \varphi_A - \alpha D_B)$, which obtains its maximum $(-\varphi_A - \alpha D_A + \alpha D_B)^2/(4\alpha D_A)$ at $(-\varphi_A + \alpha D_A + \alpha D_B)/(2\alpha D_A)$. Since $\varphi_A \leq \alpha(D_A + D_B)$, we conclude that $\rho_A^* > 0$ and because $\alpha(D_B - D_A) \leq \varphi_A$, the maximum is strictly positive, i.e., $R^{32} > R^{12}$.

- $R^{32} \geq R^{11}$: since $\varphi_B \geq \varphi_A$, $R^{32} - R^{11} \geq \frac{1}{\alpha}(\frac{\varphi_A}{2} + \frac{\alpha(D_A+D_B)}{2})^2 \geq 0$.

- $R^{32} \geq R^{21}$: since $\varphi_B \geq \varphi_A$, $R^{32} \geq \frac{1}{\alpha}(\frac{\varphi_A}{2} + \frac{\alpha(D_A+D_B)}{2})^2 + R^{11}$, and therefore $R^{32} - R^{21} \geq \frac{1}{\alpha}(\frac{\varphi_A}{2} + \frac{\alpha(-D_A+D_B)}{2})^2 \geq 0$.

- $R^{32} \geq R^{23}$: consider the region $\varphi_A \leq \varphi_B \leq \alpha(D_A + D_B)$ and let $\varphi_B = \varphi_A + \epsilon$, then we have $R^{32} - R^{23} = -\frac{\epsilon}{4\alpha}(2\varphi_B - \epsilon) + \frac{\epsilon}{2}(D_A + D_B) \geq \frac{\epsilon^2}{4\alpha} \geq 0$.

Therefore, R^{32}, i.e., the patient diversion (ρ,0) is optimal.

- Region 6: $\alpha(D_A - D_B) \leq \varphi_B \leq \alpha(D_A + D_B)$ and $\varphi_A \geq \varphi_B$.

First, we notice that S^{31} is infeasible in this region. Moreover,

- $R^{23} > R^{22}$: $R^{23} - R^{22} = D_B(-\alpha\rho_B^2 D_A + \rho_B(-\varphi_B + \alpha D_A + \alpha D_B))$, which obtains its maximum $(-\varphi_B + \alpha D_A + \alpha D_B)^2/(4\alpha D_B)$ at $(-\varphi_B + \alpha D_A + \alpha D_B)/(2\alpha D_B)$. Since $\varphi_B \leq \alpha(D_A + D_B)$, we conclude that $\rho_B^* > 0$ and the maximum is strictly positive, i.e., $R^{23} > R^{22}$.

- $R^{23} > R^{21}$: $R^{23} - R^{21} = D_B(-\alpha\rho_B^2 D_B + \rho_B(-\varphi_B + \alpha D_A + \alpha D_B) + \varphi_B - \alpha D_A)$, which obtains its maximum $(-\varphi_B - \alpha D_B + \alpha D_A)^2/(4\alpha D_B)$ at $(-\varphi_B + \alpha D_A + \alpha D_B)/(2\alpha D_B)$. Since $\varphi_B \leq \alpha(D_A + D_B)$, we conclude that $\rho_B^* > 0$ and because $\alpha(D_A - D_B) \leq \varphi_B$, the maximum is strictly positive, i.e., $R^{23} > R^{21}$.

- $R^{23} \geq R^{11}$: since $\varphi_A \geq \varphi_B$, $R^{21} - R^{33} \geq \frac{1}{\alpha}(\frac{\varphi_B}{2} + \frac{\alpha(D_A+D_B)}{2})^2 \geq 0$.

- $R^{23} \geq R^{12}$: $R^{23} - R^{12} \geq \frac{1}{\alpha}(\frac{\varphi_B}{2} + \frac{\alpha(D_A-D_B)}{2})^2 \geq 0$.

- $R^{23} \geq R^{32}$: consider the region $\varphi_B \leq \varphi_A \leq \alpha(D_A + D_B)$ and let $\varphi_A = \varphi_B + \epsilon$, then we have $R^{23} - R^{32} = -\frac{\epsilon}{4\alpha}(2\varphi_A - \epsilon) + \frac{\epsilon}{2}(D_A + D_B) \geq \frac{\epsilon^2}{4\alpha} \geq 0$.

- $R^{23} \geq R^{13}$: since $\varphi_A \geq \varphi_B$, $R^{23} \geq R^{11} + \frac{1}{\alpha}(\frac{\varphi_B}{2} + \frac{\alpha(D_A+D_B)}{2})^2 = R^{13}$.

Therefore, R^{23}, i.e., the patient diversion (0,ρ) is optimal.

- Region 7: $\alpha(-D_A - D_B) \leq \varphi_B \leq \alpha(D_B - D_A)$ and $\varphi_B \geq \varphi_A$.

First, we notice that S^{32} is infeasible in the region. Moreover, with $R^{13} = \frac{1}{\alpha}(\frac{\varphi_B}{2} + \frac{\alpha(D_A+D_B)}{2})^2 + R^{11}$, we have

- $R^{13} \geq R^{12} : R^{13} - R^{12} = D_B(-\alpha\rho_B^2 D_B + \rho_B(-\varphi_B - \alpha D_A + \alpha D_B))$, which obtains its maximum $(-\varphi_B - \alpha D_A + \alpha D_B)^2/(4\alpha D_B)$ at $(-\varphi_B - \alpha D_A + \alpha D_B)/(2\alpha D_B)$. Since $\varphi_B \leq \alpha(D_B - D_A)$, we conclude that $\rho_B^* > 0$, and the maximum is strictly positive. Thus, $R^{13} \geq R^{12}$.

- $R^{13} \geq R^{11}$: $R^{13} - R^{11} = D_B(-\alpha\rho_B^2 D_B + \rho_B(-\varphi_B - \alpha D_A + \alpha D_B) + (\varphi_B + \alpha D_A))$, which obtains its maximum $(\varphi_B + \alpha D_A + \alpha D_B)^2/(4\alpha D_B)$ at $(-\varphi_B - \alpha D_A + \alpha D_B)/(2\alpha D_B)$. Since $\varphi_B \leq \alpha(D_B - D_A)$, we conclude that $\rho_B^* > 0$, and since $\alpha(-D_A - D_B) \leq \varphi_B$ the maximum is strictly positive. Thus, $R^{13} \geq R^{11}$.

- $R^{13} \geq R^{22}$: $R^{13} \geq \varphi_B(D_A + D_B) + R^{11} \geq \varphi_A D_A + \varphi_B D_B + R^{11} = R^{22}$.

- $R^{13} \geq R^{32}$: $R^{13} \geq \frac{1}{\alpha}(\frac{\varphi_A}{2} + \frac{\alpha(D_A+D_B)}{2})^2 + R^{11} \geq D_A(\varphi_A + \alpha D_B) + R^{11} = R^{21}$.

- $R^{13} \geq R^{23}$: since $\varphi_A \leq \varphi_B$, $R^{23} \leq \frac{1}{\alpha}(\frac{\varphi_B}{2} + \frac{\alpha(D_A+D_B)}{2})^2 + R^{11} = R^{13}$.

- $R^{13} \geq R^{31}$: since $\varphi_B \geq \varphi_A$, $\frac{1}{\alpha}(\frac{\varphi_B}{2} + \frac{\alpha(D_A+D_B)}{2})^2 \geq \frac{1}{\alpha}(\frac{\varphi_A}{2} + \frac{\alpha(D_A+D_B)}{2})^2$.

Therefore, R^{13}, i.e., the patient diversion (1,ρ) is optimal.

- Region 8: Symmetric to the optimal region of S^{13} with the patient diversion $(1, \rho)$, S^{31} with $(\rho, 1)$ is optimal in the region $\alpha(-D_A - D_B) \leq \varphi_A \leq \alpha(D_A - D_B)$ and $\varphi_B \leq \varphi_A$.

Non-Zero Set-up Cost.

- When $\theta < \alpha D_A^2$, the partition is shown in Figure 3(b).

On x-axis:

Partition line between S^{11} and S^{31}: $\frac{-D_A-D_B}{2D_A} + \frac{\sqrt{\theta}}{\sqrt{\alpha}D_A}$;

Partition line between S^{31} and S^{21}: $\frac{D_A-D_B}{2D_A}$;

Partition line between S^{12} and S^{32}: $\frac{D_B-D_A}{2D_A} + \frac{\sqrt{\theta}}{\sqrt{\alpha}D_A}$;

Partition line between S^{32} and S^{22}: $\frac{D_A+D_B}{2D_A}$.

On y-axis:

Partition line between S^{11} and S^{13}: $\frac{-D_A-D_B}{2D_B} + \frac{\sqrt{\theta}}{\sqrt{\alpha}D_B}$;

Partition line between S^{13} and S^{12}: $\frac{D_B-D_A}{2D_B}$;

Partition line between S^{21} and S^{23}: $\frac{D_A-D_B}{2D_B} + \frac{\sqrt{\theta}}{\sqrt{\alpha}D_B}$;

Partition line between S^{23} and S^{22}: $\frac{D_A+D_B}{2D_B}$.

The partition line between S^{13} and S^{21} is a **parabola** $\alpha D_B^2\zeta_B^2 + \alpha D_B(D_A + D_B)\zeta_B - 2\alpha D_A^2\zeta_A + \frac{\alpha}{4}(D_A - D_B)^2 = 0$.

The partition line between S^{12} and S^{23} is a **parabola** $-\alpha D_B^2\zeta_B^2 + \alpha D_B(D_A + D_B)\zeta_B - 2\alpha D_A^2\zeta_A - \frac{\alpha}{4}(D_A - D_B)^2 + \theta = 0$.

The partition line between S^{13} and S^{23} is a straight line with slope $\frac{D_A}{D_B}$: $-2\alpha D_A^2\zeta_A + 2\alpha D_A D_B\zeta_B + \theta = 0$.

The slope for the partition line between S^{31} and S^{13} is $\frac{D_A}{D_B}$.

- When $\alpha D_A^2 < \theta < \alpha D_B^2$, the partition is shown in Figure 3(c).

On x-axis:

Partition line between S^{11} and S^{21}: $\frac{-D_B}{2D_A} + \frac{\theta}{2\alpha D_A^2}$;

Partition line between S^{12} and S^{22}: $\frac{D_B}{2D_A} + \frac{\theta}{2\alpha D_A^2}$.

On y-axis:

Partition line between S^{11} and S^{13}: $\frac{-D_A - D_B}{2D_B} + \frac{\sqrt{\theta}}{\sqrt{\alpha} D_B}$;

Partition line between S^{13} and S^{12}: $\frac{D_B - D_A}{2D_B}$;

Partition line between S^{21} and S^{23}: $\frac{D_A - D_B}{2D_B} + \frac{\sqrt{\theta}}{\sqrt{\alpha} D_B}$;

Partition line between S^{23} and S^{22}: $\frac{D_A + D_B}{2D_B}$.

- When $\alpha D_B^2 < \theta$, the partition is shown in Figure 3(d).

On x-axis:

Partition line between S^{11} and S^{21}: $\frac{-D_B}{2D_A} + \frac{\theta}{2\alpha D_A^2}$;

Partition line between S^{12} and S^{22}: $\frac{D_B}{2D_A} + \frac{\theta}{2\alpha D_A^2}$.

On y-axis:

Partition line between S^{11} and S^{12}: $\frac{-D_A}{2D_B} + \frac{\theta}{2\alpha D_B^2}$;

Partition line between S^{21} and S^{22}: $\frac{D_A}{2D_B} + \frac{\theta}{2\alpha D_B^2}$.

The slope for the partition line between S^{21} and S^{12} is $\frac{D_A^2}{D_B^2}$.

The partition line in Proposition 3 for a non-zero θ is derived similarly with modification of the formulae to capture differences between contiguous regions. We omit the detailed proofs here. ∎

Proof of Theorem 1

Proof. We rank the communities according to their φ_i's in an ascending order. Without loss of generality, assume we have $|I| = N$ ordered communities with unique φ_i's. First, we define that excluding k, the congestion at the hospital is $W_k^C = \sum_{j\neq k} \rho_j D_j$ and the total number of patients staying at home is $H_k^C = \sum_{j\neq k}(1-\rho_j)D_j$. With a universal linear nurse coordination cost $g(D) = \beta D$, we further write the objective function as

$$\begin{aligned}R = &(1-\rho_k)D_k(-\Delta\mathcal{R} + f_k + \alpha(\rho_k D_k + W_k^C))\\ &+\sum_{j\neq k}(1-\rho_j)D_j(-\Delta\mathcal{R} + f_j + \alpha(\rho_k D_k + W_k^C))\\ &-\gamma(\rho_k D_k + W_k^C) - \beta(H_k^C + (1-\rho_k)D_k).\end{aligned}$$

This is true for any $\rho_k \in [0,1]$, $k \in I$, and the Lagrange dual is $L = R + \lambda_k\rho_k + \mu_k(1-\rho_k)$, where $\lambda_k \geq 0$ and $\mu_k \geq 0$. Then, we can calculate the optimal value of ρ_k given that W_k^C and H_k^C are known following the FOC:

$$\frac{\partial L}{\partial \rho_k} = -(\varphi_k + \alpha\rho_k D_k + \alpha W_k^C)D_k + \alpha(1-\rho_k)D_k^2 + \alpha H_k^C D_k + \lambda_k - \mu_k = 0,$$

which yields $\rho_k^* = \max\{0, \min\{1, \frac{D_k + H_k^C}{2D_k} - \frac{\varphi_k + \alpha W_k^C}{2\alpha D_k}\}\}$, where $\varphi_k = f_k + \gamma - \beta$.

After calculating all ρ_k^*'s, $k \in I$, we can further check if they are the genuine optimal solution to the original problem. If they do, the following two scenarios can unfold.

1. There exists at least one k such that $\rho_k^* \in (0,1)$. Under this scenario, we further consider, if $\exists\ j,k,\ s.t.\ \rho_j^*, \rho_k^* \in (0,1)$. Suppose both ρ_j^* and ρ_k^* are the genuine optimal solution to the optimization problem, with the optimal congestion W^* (the total number of patients going to the hospital under the optimal solution) and the optimal total number of patients staying at home H^*. Then, we have $W_k^C = W^* - \rho_k^* D_k$ and $H_k^C = H^* - (1-\rho_k^*)D_k$, and $\rho_k^* = \frac{D_k + H^* - (1-\rho_k^*)D_k}{2D_k} - \frac{\varphi_k + \alpha(W^* - \rho_k^* D_k)}{2\alpha D_k}$, which suggests that $\alpha(H^* - W^*) = \varphi_k$. Similarly, we have $\alpha(H^* - W^*) = \varphi_j$. Since φ_j and φ_k are unequal according to our assumption, the optimal solution does not allow more than one community, denoted as k, to have $\rho_k^* \in (0,1)$. Furthermore, since the genuine optimal ρ_k^* follows $\rho_k^* = \max\{0, \min\{1, \frac{D_k + H^* - (1-\rho_k^*)D_k}{2D_k} - \frac{\varphi_k + \alpha(W^* - \rho_k^* D_k)}{2\alpha D_k}\}\}$, then, for all $\varphi_i < \varphi_k$, we have $\alpha(H^* - W^*) > \varphi_i$, and thus $\rho_i^* = 1$. For all $\varphi_i > \varphi_k$, $\alpha(H^* - W^*) < \varphi_i$, and $\rho_i^* = 0$.
2. Else, $\rho_k^* \in \{0,1\}$, $\forall k \in I$:

—If $\rho_k^* = 1$, $\rho_k^* \leq \frac{D_k + H^* - (1-\rho_k^*)D_k}{2D_k} - \frac{\varphi_k + \alpha(W^* - \rho_k^* D_k)}{2\alpha D_k}$, which suggests $\alpha(H^* - W^*) - \varphi_k \geq 0$. Then, for $i < k$ with $\varphi_i < \varphi_k$, $\rho_i^* = 1$, and $\alpha(H^* - W^*) - \varphi_i \geq 0$.

—If $\rho_k^* = 0$, $\rho_k^* \geq \frac{D_k + H^* - (1-\rho_k^*)D_k}{2D_k} - \frac{\varphi_k + \alpha(W^* - \rho_k^* D_k)}{2\alpha D_k}$, which suggests $\alpha(H^* - W^*) - \varphi_k \leq 0$. Then, for $i > k$ with $\varphi_i > \varphi_k$, $\rho_i^* = 0$, and $\alpha(H^* - W^*) - \varphi_i \leq 0$.

Collectively, for the genuine optimal solution, there exists a threshold $K \in I$, for $i < K$, $\rho_i^* = 1$ and for $i > K$, $\rho_i^* = 0$. Since $H^* + W^* = \sum_j D_j$, when $\rho_K^* \in (0,1)$, it naturally follows that $W^* = \frac{\alpha \sum_j D_j - \varphi_K}{2\alpha}$. More generally, based on the FOC:

$$\rho_i^* \in (0,1) \text{ if and only if } W^* = \frac{\alpha \sum_j D_j - \varphi_i}{2\alpha};$$
$$\rho_i^* = 1 \text{ if and only if } W^* \leq \frac{\alpha \sum_j D_j - \varphi_i}{2\alpha};$$
$$\rho_i^* = 0 \text{ if and only if } W^* \geq \frac{\alpha \sum_j D_j - \varphi_i}{2\alpha}.$$

■

Given that the genuine optimal solution has the structure of having one "threshold" community, we further introduce the following notations that enable us to pin down the threshold community, which is the key to the optimization problem. Define $\rho_k^{<m>}$ as the optimal partition of community k when community m serves as the candidate threshold community, and similarly, define $W_k^{<m>,C}$ ($H_k^{<m>,C}$), the number of patients going to the hospital (staying at home, respectively) except for community k patients when community m serves as the candidate threshold community. In addition, let $R^{<m>}$ be the revenue when community m serves as the candidate threshold community. Based on Theorem 1, $\rho_k^* = \max\{0, \min\{1, \frac{D_k + H_k^C}{2D_k} - \frac{\varphi_k + \alpha W_k^C}{2\alpha D_k}\}\}$, given that W_k^C and H_k^C are known. Then, we come up with the following lemma.

Lemma A.2. *The optimal solution given a candidate threshold community has a monotonic property:*

(i) If there exists $\rho_k^{<k>} = \frac{D_k + H_k^{<k>,C}}{2D_k} - \frac{\varphi_k + \alpha W_k^{<k>,C}}{2\alpha D_k} \in [0,1)$, *then* $\rho_{k+1}^{<k+1>} = 0$.

(ii) If there exists $\rho_k^{<k>} = \frac{D_k + H_k^{<k>,C}}{2D_k} - \frac{\varphi_k + \alpha W_k^{<k>,C}}{2\alpha D_k} \in (0,1]$, *then* $\rho_{k-1}^{<k-1>} = 1$.

(iii) $\rho_k^{<k>}$ *is monotonically decreasing with* k.

Proof. First, if there exists $\rho_k^{<k>} = \frac{D_k + H_k^{<k>,C}}{2D_k} - \frac{\varphi_k + \alpha W_k^{<k>,C}}{2\alpha D_k} \in [0,1)$, our goal is to prove $\rho_{k+1}^{<k+1>} = 0$. Note that $\rho_k^{<k>} < 1$, so $\frac{H_k^{<k>,C}}{2} - \frac{\varphi_k + \alpha W_k^{<k>,C}}{2\alpha} < \frac{D_k}{2}$. Furthermore, since $H_{k+1}^{<k+1>,C} = H_k^{<k>,C} - D_{k+1}$, $W_{k+1}^{<k+1>,C} = W_k^{<k>,C} + D_k$, (see Figure A.1 for illustration), and $\varphi_{k+1} > \varphi_k$, we have $\rho_{k+1}^{<k+1>} = \max\{0, \min\{\rho_{k+1}, 1\}\}$, where

$$\rho_{k+1} = \frac{D_{k+1} + H_{k+1}^{<k+1>,C}}{2D_{k+1}} - \frac{\varphi_{k+1} + \alpha W_{k+1}^{<k+1>,C}}{2\alpha D_{k+1}}$$

$$
\begin{aligned}
&= \frac{D_{k+1} + H_k^{<k>,C} - D_{k+1}}{2D_{k+1}} - \frac{\varphi_{k+1} + \alpha W_k^{<k>,C} + \alpha D_k}{2\alpha D_{k+1}} \\
&< \frac{H_k^{<k>,C}}{2D_{k+1}} - \frac{\varphi_k + \alpha W_k^{<k>,C}}{2\alpha D_{k+1}} - \frac{D_k}{2D_{k+1}} \\
&< \frac{D_k}{2D_{k+1}} - \frac{D_k}{2D_{k+1}} = 0.
\end{aligned}
$$

Therefore, $\rho_{k+1}^{<k+1>} = 0$.

Similarly, if there exists $\rho_k^{<k>} = \frac{D_k + H_k^{<k>,C}}{2D_k} - \frac{\varphi_k + \alpha W_k^{<k>,C}}{2\alpha D_k} \in [0,1)$, we aim to prove $\rho_{k-1}^{<k-1>} = 1$. With $\rho_k^{<k>} > 0$, $\frac{H_k^{<k>,C}}{2} - \frac{\varphi_k + \alpha W_k^{<k>,C}}{2} > -\frac{D_k}{2}$. Since $H_{k-1}^{<k-1>,C} = H_k^{<k>,C} + D_k$, $W_{k-1}^{<k-1>,C} = W_k^{<k>,C} - D_{k-1}$, and $\varphi_{k-1} < \varphi_k$, we have $\rho_{k-1}^{<k-1>} = \max\{0, \min\{\rho_{k-1}, 1\}\}$, where

$$
\begin{aligned}
\rho_{k-1} &= \frac{D_{k-1} + H_{k-1}^{<k-1>,C}}{2D_{k-1}} - \frac{\varphi_{k-1} + \alpha W_{k-1}^{<k-1>,C}}{2\alpha D_{k-1}} \\
&= \frac{D_{k-1} + H_k^{<k>,C} + D_k}{2D_{k-1}} - \frac{\varphi_{k-1} + \alpha W_k^{<k>,C} - \alpha D_{k-1}}{2\alpha D_{k-1}} \\
&> \frac{1}{2} + \frac{H_k^{<k>,C}}{2D_{k-1}} - \frac{\varphi_k + \alpha W_k^{<k>,C}}{2\alpha D_{k-1}} + \frac{D_k}{2D_{k-1}} + \frac{1}{2} \\
&> 1 - \frac{D_k}{2D_{k-1}} + \frac{D_k}{2D_{k-1}} = 1.
\end{aligned}
$$

Therefore, $\rho_{k-1}^{<k-1>} = 1$. The monotonic property is a direct result following (i) and (ii). ■

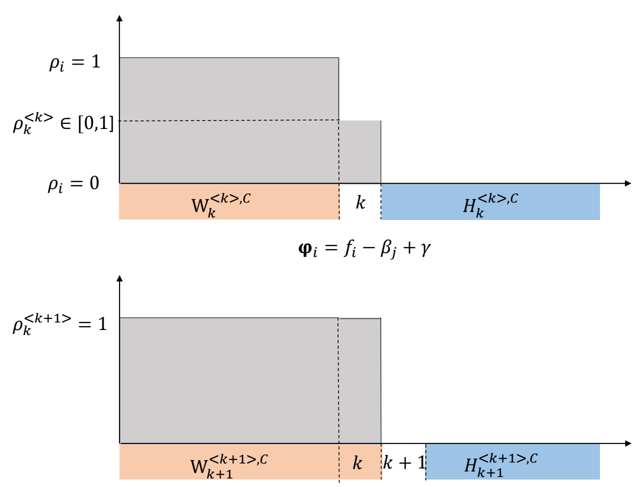


Figure A.1 An illustration of the relationship between $W_k^{<k>,C}$ and $H_k^{<k>,C}$.

Lemma A.3. *The revenue obtained when an arbitrary community m is selected as the candidate threshold community is defined as $R^{<m>}$. It has the following monotonic property:*

(i) For $1 < k \leq N$*, if* $R^{<k>} > R^{<k-1>}$*, then* $R^{<k-1>} > R^{<k-2>}$*.*

(ii) For $1 \leq k < N-1$*, if* $R^{<k>} > R^{<k+1>}$*, then* $R^{<k+1>} > R^{<k+2>}$*.*

Proof. We start with the proof of (i) if $R^{<k>} > R^{<k-1>}$, then, $R^{<k-1>} > R^{<k-2>}$. Based on the definition, two scenarios can unfold.

First, $\rho_k^{<k>} \in (0,1)$. Based on Lemma A.2, $\rho_{k-1}^{<k-1>} = 1$, and based on Theorem 1, $\rho_k^{<k-1>} = 0$. Let $W_k^{<k>,C} = \sum_{j=1}^{k-1} D_j$, and $H_k^{<k>,C} = \sum_{j=k+1}^{N} D_j$. We compare the revenue when k or $k-1$ is selected as the candidate threshold community.

$$\begin{aligned} R^{<k>} - R^{<k-1>} =& \alpha H_k^{<k>,C} \rho_k^{<k>} D_k - D_k(\varphi_k + \alpha W_k^{<k>,C}) \\ &+ D_k(1-\rho_k^{<k>})(\varphi_k + \alpha W_k^{<k>,C} + \alpha \rho_k^{<k>} D_k) \\ =& \alpha \rho_k^{<k>} D_k H_k^{<k>,C} - \rho_k^{<k>} D_k(\varphi_k + \alpha W_k^{<k>,C}) + \alpha(1-\rho_k^{<k>})\rho_k^{<k>} D_k^2. \end{aligned}$$

The assumption $R^{<k>} - R^{<k-1>} > 0$ is equivalent to

$$\alpha \rho_k^{<k>} D_k H_k^{<k>,C} - \rho_k^{<k>} D_k(\varphi_k + \alpha W_k^{<k>,C}) + \alpha(1-\rho_k^{<k>})\rho_k^{<k>} D_k^2 > 0.$$

Since $\rho_k^{<k>} > 0$, this condition implies $\alpha H_k^{<k>,C} - (\varphi_k + \alpha W_k^{<k>,C}) + \alpha(1-\rho_k^{<k>}) D_k > 0$.

Based on Lemma A.2, $\rho_{k-1}^{<k-1>} = 1$, and then $\rho_{k-2}^{<k-2>} = 1$, and based on Theorem 1, $\rho_{k-2}^{<k-1>} = 0$. Then, one can calculate the difference between $R^{<k-1>}$ and $R^{<k-2>}$:

$$\begin{aligned} R^{<k-1>} - R^{<k-2>} =& \alpha(D_k + H_k^{<k>,C}) D_{k-1} - D_{k-1}(\varphi_{k-1} + \alpha W_k^{<k>,C} - \alpha D_{k-1}) \\ >& D_{k-1}(\alpha H_k^{<k>,C} - (\varphi_k + \alpha W_k^{<k>,C}) + \alpha D_k + \alpha D_{k-1}) > 0. \end{aligned}$$

Next, in the case of $\rho_k^{<k>} = 1$. Using the same argument, $\rho_k^{<k-1>} = 0$. Then, we compare the difference between $R^{<k>}$ and $R^{<k-1>}$.

$$R^{<k>} - R^{<k-1>} = \alpha D_k H_k^{<k>,C} - D_k(\varphi_k + \alpha W_k^{<k>,C}) > D_k(\alpha H_k^{<k>,C} - (\varphi_k + \alpha W_k^{<k>,C})).$$

The assumption $R^{<k>} - R^{<k-1>} > 0$ is equivalent to $\alpha H_k^{<k>,C} - (\varphi_k + \alpha W_k^{<k>,C}) > 0$. Then,

$$R^{<k-1>} - R^{<k-2>} = \frac{D_{k-1}}{D_k}(R^{<k>} - R^{<k-1>}) + \alpha D_{k-1}(D_k + \alpha D_{k-1}) > 0.$$

The proof of (ii) if $R^{<k>} > R^{<k+1>}$, then $R^{<k+1>} > R^{<k+2>}$ can be done analogously, and we omit the derivations. ■

Proof of Proposition 2

Proof. To obtain the threshold community K and ρ_K^*, we first calculate $W_k^{<k>,C} = \sum_{j=1}^{k-1} D_j$, and $H_k^{<k>,C} = \sum_{j=k+1}^{N} D_j$, and obtain $\rho_k^{<k>} = \max\{0, \min\{1, \frac{D_k + H_k^{<k>,C}}{2D_k} - \frac{\varphi_k + \alpha W_k^{<k>,C}}{2\alpha D_k}\}\}$, for each $k = 1...N$, and the corresponding revenue is denoted as $R^{<k>}$. The true threshold community satisfies $K = \text{argmax}_k R^{<k>}$.

Further, let $\rho_k = \frac{D_k + H_k^{<k>,C}}{2D_k} - \frac{\varphi_k + \alpha W_k^{<k>,C}}{2\alpha D_k}$. If $\rho_k \in (0,1)$, then, $\alpha(H_k^{<k>,C} - W_k^{<k>,C}) - \varphi_k = 2\alpha\rho_k D_k - \alpha D_k$.

$$\begin{aligned}
\text{For} \quad 1 < k \leq N, \quad R^{<k>} - R^{<k-1>} &= \rho_k D_k(\alpha(H_k^{<k>,C} - W_k^{<k>,C}) - \varphi_k + \alpha(1-\rho_k)D_k) \\
&= \rho_k D_k((1-\rho_k)\alpha D_k + 2\alpha\rho_k D_k - \alpha D_k) \\
&= \alpha\rho_k^2 D_k^2 > 0. \\
\text{For} \quad 1 \leq k < N, \quad R^{<k+1>} - R^{<k>} &= (1-\rho_k)D_k(\alpha(H_k^{<k>,C} - W_k^{<k>,C}) - \varphi_k - \alpha\rho_k D_k) \\
&= -\alpha(1-\rho_k)^2 D_k^2 < 0.
\end{aligned}$$

Moreover, if $\rho_k \geq 1$, then, $\alpha(H_k^{<k>,C} - W_k^{<k>,C}) - \varphi_k \geq \alpha D_k > 0$. In addition, $\rho_k^{<k-1>} = 0$, according to Theorem 1 and Lemma A.2. Then,

$$R^{<k>} - R^{<k-1>} = D_k(\alpha(H_k^{<k>,C} - W_k^{<k>,C}) - \varphi_k) > 0.$$

If $\rho_k \leq 0$, then, $\alpha(H_k^{<k>,C} - W_k^{<k>,C}) - \varphi_k \leq -\alpha D_k < 0$. With the same argument, $\rho_k^{<k+1>} = 1$. Then,

$$R^{<k+1>} - R^{<k>} = D_k(\alpha(H_k^{<k>,C} - W_k^{<k>,C}) - \varphi_k) < 0.$$

Since we have $\rho_k^{<k>} = \max\{0, \min\{\rho_k, 1\}\}$, we can conclude that if $\rho_k^{<k>} \in (0,1)$, then, k is the true threshold community, i.e., $K = k$ and $R^{<k>}$ is the optimal revenue. If $\rho_k^{<k>} = 1$, then $R^{<k>} > R^{<k-1>}$; if $\rho_k^{<k>} = 0$, then $R^{<k>} > R^{<k+1>}$. ∎

Corollary A.1. *If $\forall i$, $\varphi_i \geq \alpha \sum_i D_i$, then $\rho_i = 0$, i.e., all patients should stay at home.*

Proof of Proposition 3

Proof. Proposition 2 together with Lemma A.3 suggest that $R^{<i>}$ is a unimodal function of i, and thus Algorithm 1 reduces to an $O(log(N))$ algorithm. ∎

Proof of Theorem 2

Proof. Assume the optimal solution recommends two communities to adopt mixed strategies, denoted as community A and community B. Let $d_A^* = (1-\rho_A^*)D_A$ and $d_B^* = (1-\rho_B^*)D_B$ be the optimal stay-at-home demands, and we consider a positive change $\epsilon > 0$. Denote superscript "1" to indicate the results obtained with $d_A^1 = d_A^* - \epsilon$ and superscript "2" as that of $d_A^2 = d_A^* + \epsilon$. The corresponding nurse costs are g_A^1 and g_A^2. Furthermore, let $g_A^* = g_A(d_A^*)$. The same notation applies for community B. Because $g_i(D)$'s are non-decreasing concave functions, $2g_A^* - g_A^1 - g_A^2 \geq 0$ and $2g_B^* - g_B^1 - g_B^2 \geq 0$. See Figure A.2 for illustration. Since d_A^* and d_B^* are optimal, it follows that

$$
\begin{aligned}
&d_A^*(f_A - \Delta\mathcal{R} + \gamma + W^*) - g_A(d_A^*) + d_B^*(f_B - \Delta\mathcal{R} + \gamma + W^*) - g_B(d_B^*) \\
\geq &(d_A^* - \epsilon)(f_A - \Delta\mathcal{R} + \gamma + W^*) - g_A(d_A^* - \epsilon) + (d_B^* + \epsilon)(f_B - \Delta\mathcal{R} + \gamma + W^*) - g_B(d_B^* + \epsilon),
\end{aligned}
$$

i.e.,

$$
0 \geq \epsilon(f_B - f_A) + (g_A^* - g_A^1) + (g_B^* - g_B^2). \tag{A.5}
$$

Similarly, because of the symmetry of A and B, with the same argument, we know

$$
0 \geq \epsilon(f_A - f_B) + (g_B^* - g_B^1) + (g_A^* - g_A^2). \tag{A.6}
$$

Combining the two inequalities, we get $0 \geq (2g_A^* - g_A^1 - g_A^2) + (2g_B^* - g_B^2 - g_B^1) \geq 0$.

This is true if and only if the cost functions are linear in their own domain $[d_i^* - \epsilon, d_i^* + \epsilon]$, for $i \in \{A, B\}$. Define β_i as the gradient of the cost function $g_i(D)$ when D is in $[d_i^* - \epsilon, d_i^* + \epsilon]$, based on which, we have $g_A^1 + 2\beta_A\epsilon = g_A^* + \beta_A\epsilon = g_A^2$ and $g_B^1 + 2\beta_B\epsilon = g_B^* + \beta_B\epsilon = g_B^2$.

Consequently, inequalities (A.5) and (A.6) become $0 \geq \epsilon(f_B - f_A + (\beta_A - \beta_B))$ and $0 \geq \epsilon(f_A - f_B + (\beta_B - \beta_A))$. Since $\varphi_i = f_i - \beta_i + \gamma - \Delta\mathcal{R}$, $i \in \{A, B\}$, it implies $\varphi_A = \varphi_B$. In this case, there can be multiple optimal solutions that lead to the same revenue, similar to the two-community case with $\varphi_A = \varphi_B$ that as long as the total amount of patients coming to hospital is fixed, the distribution to the two communities can be flexible. Take $\epsilon = \min\{d_A^*, D_A - d_A^*, d_B^*, D_B - d_B^*\}$, then, if either function $g_i(D)$ becomes nonlinear in $[d_i^* - \epsilon, d_i^* + \epsilon]$, the two inequalities cannot hold, or otherwise, if both keep being linear, and thus the boundary such as 0 or D_i is enclosed in $[d_i^* - \epsilon, d_i^* + \epsilon]$, then, one can push at least one community to the domain boundary and induce a different equilibrium that only allows one community to adopt a mixed strategy. Therefore, there always exists an optimal solution such that at most one community $K \in I$ has $\rho_K \in (0,1)$, and all other communities have patient diversions $\rho_i = 0$ or 1, $i \neq K$. ■

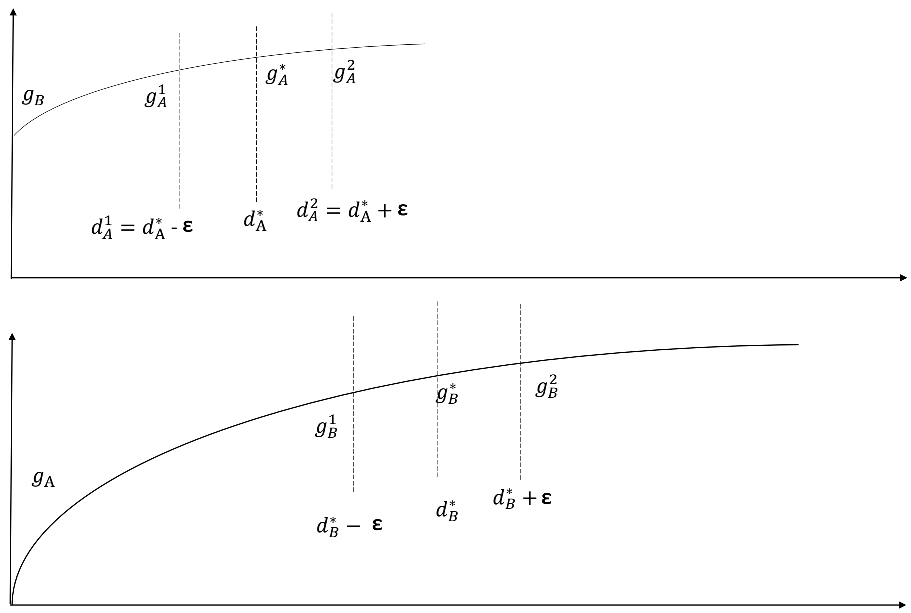


Figure A.2 Illustration of the marginal utility gain as a function of input demand.

Proof of Theorem 3

Proof. For brevity, we introduce a vector $\boldsymbol{\rho}$ as the patient diversion of all communities. With a general concave cost structure, the problem is NP-hard and the optimal solution $\boldsymbol{\rho}^{\text{gen}}$ is difficult to solve. In the following, we show that the optimal solution under the linear cost assumption $\boldsymbol{\rho}^{\text{lin}}$ is a good approximation.

We linearize any non-decreasing concave cost function $g_i(D)$ (a function of demand D) for a community i to be $g_i^{lin}(D) = \beta_i D$. The slope is obtained as $\beta_i = g_i(D_i)/D_i$ (see Figure A.3). The choice of the slope enables $g_i(0) = g_i^{lin}(0) = 0$ and $g_i(D_i) = g_i^{lin}(D_i)$. Because for general concave cost functions, there exists an optimal solution that dictates at most one community to have a partial diversion, we would expect the impact of a linear function compared to a concave one to the objective function would only present in one community, which could be small and hence leads to a good approximation.

First, we claim that $\boldsymbol{\rho}^{\text{lin}}$ is a feasible solution, because the cost function only impacts the objective function, but not the feasibility of any solutions, since the IC constraints only affect the choice of the prices $C_{2,i}$. Thus, $R^G(\boldsymbol{\rho}^{\text{lin}})$ is a lower bound, i.e.,

$$R^G(\boldsymbol{\rho}^{\text{lin}}) \leq R^G(\boldsymbol{\rho}^{\text{gen}}).$$

Next, we prove that $R^{\text{lin}}(\boldsymbol{\rho}^{\text{lin}})$ is an upper bound. We therefore show that

$$R^G(\boldsymbol{\rho}^{\text{gen}}) \leq R^{\text{lin}}(\boldsymbol{\rho}^{\text{gen}}) \leq R^{lin}(\boldsymbol{\rho}^{\text{lin}}).$$

The first inequality holds because of the property of concave functions: given solution $\boldsymbol{\rho}^{gen}$, $g((1-\rho_i)D_i) \geq \beta_i(1-\rho_i)D_i$, and the other terms in the revenue gain remain the same; the second one holds because $\boldsymbol{\rho}^{\text{lin}}$ maximizes R^{lin}.

Then, we show that the difference between the upper bound and the lower bound is smaller than $g_K(D_K)$, where K is the threshold community identified using the linear approximation, i.e.,

$$R^{\text{lin}}(\boldsymbol{\rho}^{\text{lin}}) - R^G(\boldsymbol{\rho}^{\text{lin}}) = g_K((1-\rho_K)D_K) - \beta_K(1-\rho_K)D_K \leq g_K(D_K).$$

The last inequality holds because we know when $\rho_K = 0$ or 1, the difference is zero, and the difference achieves its maximum when the increasing concave function is close to a rectangular shape (g_i^{worst}), as shown in Figure A.3.

As a remark, it is not necessary that the same community is partially diverted under the general concave cost function and the linear case, nor would the assignment of other communities stay the same. However, this does not affect the bound of the approximation gap. ■

Proof of Proposition 4

Proof. This proof follows the proof of a general Minorize-Maximization problem (a special case of an Expectation-Maximization problem). In each iteration, a "new" objective function is defined consisting of the original objective function and a penalty term, which penalizes the difference between an intermediate variable and its "expected" value obtained from the previous iteration. In the "optimization" step, such an expected value will be treated as an input, and a new set of optimal solutions will be obtained accordingly. Then, in the "expectation" step, we update the expected value of that intermediate variable using the optimal solutions obtained in the previous

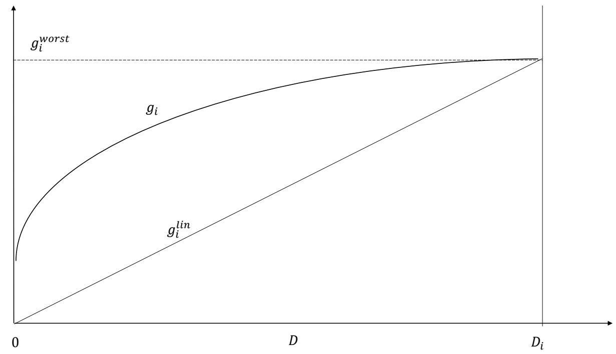


Figure A.3 Illustration of the linear approximation to the concave function.

optimization step. The new objective function and thus the original objective function will have to be non-decreasing over iterations, and then, such a process will converge to the local optima, according to (Hunter and Lange 2000). In the following we elaborate the algorithm and show how it conforms to the Minorize-Maximization problem.

1. Initialization $t=0$. We introduce $H_{t,i}$ ($W_{t,i}$) as the amount of stay-home (go-to-hospital, respectively) patients at iteration t after updating patient diversion ρ_i^t at the i-th community. One can directly initialize the algorithm with $H_{0,N}=0$ (stay-home) and $W_{0,N}=\sum_i D_i$ (go-to-hospital), and $\rho_i^0=1,\ i\in I$. An alternative initialization is to use the approximated solution obtained in Algorithm 1 by assuming $\beta_i=g(D_i)/D_i$.

2. Update $t=t+1$.

For each community i,

If $i=1$, $H_{t,i}^C=H_{t-1,N}-(1-\rho_i^{t-1})D_i$, $W_{t,i}^C=W_{t-1,N}-\rho_i^{t-1}D_i$.

If $i=2...N$, $H_{t,i}^C=H_{t,i-1}-(1-\rho_i^{t-1})D_i$, $W_{t,i}^C=W_{t,i-1}-\rho_i^{t-1}D_i$.

Define $p^{t,i}=(W_{t,i}^C-\sum_{j<i}\rho_j^tD_j-\sum_{i<j}\rho_j^{t-1}D_j)^2\geq 0$ as a penalty term, which penalizes the difference between an intermediate variable and its "expected" value obtained from the previous iteration, and introduce a new objective function $R^{t,i}(\cdot)$ described below. Then, ${\rho_i^t}^*$ is further introduced as

$$\begin{aligned}{\rho_i^t}^* &= \text{argmax}_{\rho_i} R^{t,i}(\{\rho_j^t\}_{j<i},\rho_i,\{\rho_j^{t-1}\}_{j>i})\\ &:= R(\{\rho_j^t\}_{j<i},\rho_i,\{\rho_j^{t-1}\}_{j>i})-(W_{t,i}^C-\sum_{j<i}\rho_j^tD_j-\sum_{i<j}\rho_j^{t-1}D_j)^2.\end{aligned}$$

Based on the proof of Theorem 1, $\rho_i^{t^*} = \max\{0, \min\{1, \frac{D_i + H_{t,i}^C}{2D_i} - \frac{\varphi_i^* + \alpha W_{t,i}^C}{2\alpha D_i}\}\}$, where φ_i^* is associated with the gradient of $g_i(D)$ at $D = (1-\rho_i^{t^*})D_i$. This implies solving for an implicit function, and $\rho_i^{t^*}$ cannot be directly obtained.

—Here assume there are $L_i \geq 1$ segments for each piece-wise linear concave function $g_i^{\text{p-lin}}(D)$ as an approximation to $g_i(D)$, and let β_i^l be the gradient of the cost function at the l^{th} segment, $1 \leq l \leq L_i$. Let $\varphi_i^l = f_i + \gamma - \beta_i^l - \Delta\mathcal{R}$ for $1 \leq l \leq L_i$. In addition, introduce $\rho_i^{t,l}$ as the optimal diversion of community i at iteration t by assuming the gradient of the cost function falls on the l^{th} segment with value β_i^l. Define $R^{t,l} = R(\rho_i^{t,l})$ as the optimal revenue when community i adopts the optimal solution $\rho_i^{t,l}$ and the other communities keep their current diversions, i.e., ρ_j^t for $j < i$ and ρ_j^{t-1} for $j > i$. We deem a solution $\rho_i^{t,l}$ as feasible if $(1-\rho_i^{t,l})D_i$ falls on the domain of segment l; otherwise, it cannot serve as a candidate of the optimal solution. Furthermore, for the cost function with a non-zero set-up cost, denote $\rho_i^t = 1$ (implying no nurse coordination cost) as $\rho_i^{t,l=0}$, and the corresponding revenue as $R^{t,l=0}$. Then, we define the optimal solution of ρ_i at iteration t as $\rho_i^{t^*} = \operatorname{argmax}_{\rho_i^{t,l}} R^{t,l}$, for $0 \leq l \leq L_i$.

—We hereby show that at least one $\rho_i^{t,l}$ is feasible, in the sense that it is a fixed point to the implicit function. Because of being concave, $\beta_i^l < \beta_i^{l-1}$ and consequently $\varphi_i^l < \varphi_i^{l-1}$. Meanwhile, because $\rho_i^{t,l} = \max\{0, \min\{1, \frac{D_i + H_{t,i}^C}{2D_i} - \frac{\varphi_i^l + \alpha W_{t,i}^C}{2\alpha D_i}\}\}$, we know that $\rho_i^{t,l}$ is monotone decreasing with φ_i^l. Therefore, $\rho_i^{t,l} < \rho_i^{t,l-1}$. For each segment, define $h_i^{t,l}$ as the amount of patients staying in the community given $\rho_i^{t,l}$, and $h_i^{t,l}$ is monotone increasing with φ_i^l. The two end points for each segment l are further denoted as h_{l-1}^U and h_l^U, and we omit i here for brevity. Assume no $\rho_i^{t,l}$ is feasible, then we prove by induction that $h_i^{t,l} > h_l^U$, for $1 \leq l < L$. First, for the first piece with a slope with β_1 and two endpoints 0 and h_1^U, the infeasibility of $\rho_i^{t,1}$ leads to $h_i^{t,1} > h_1^U$. Then, assume $h_i^{t,l-1} > h_{l-1}^U$ is true. Move to the case of l, the assumption $\rho_i^{t,l}$ is infeasible leads to either $h_i^{t,l} < h_{l-1}^U$ or $h_i^{t,l} > h_l^U$; however, with $h_i^{t,l} > h_i^{t,l-1}$ (monotonic property), and $h_i^{t,l-1} > h_{l-1}^U$ we know that $h_i^{t,l} > h_{l-1}^U$, and therefore, it can only be $h_i^{t,l} > h_l^U$. Lastly, the assumption $\rho_i^{t,L}$ is infeasible leads to $h_i^{t,L} > h_L^U = D_i$. However, one cannot obtain a solution that is out of the domain bound. Therefore, a conflict is induced and thus, there must be one $\rho_i^{t,l}$ being feasible.

Let $\rho_i^t = \rho_i^{t*}$ and update $W_{t,i} = \sum_{j\le i} \rho_j^t D_j + \sum_{i<j} \rho_j^{t-1} D_j$ and $H_{t,i} = \sum_{j\le i}(1-\rho_j^t)D_j + \sum_{i<j}(1-\rho_j^{t-1})D_j$. Based on the definition that ρ_i^t optimizes $R^{t,i}(\cdot)$,

$$\begin{aligned} &R^{t,i}(\{\rho_j^t\}_{j<i}, \rho_i^t, \{\rho_j^{t-1}\}_{j>i}) \ge R^{t,i}(\{\rho_j^t\}_{j<i}, \rho_i^{t-1}, \{\rho_j^{t-1}\}_{j>i}) \\ =&R^{t,i}(\{\rho_j^t\}_{j<i-1}, \rho_{i-1}^t, \{\rho_j^{t-1}\}_{j>i-1}) = \max R^{t,i-1}(\cdot). \end{aligned}$$

Thus, the objective function $R^{t,i}(\cdot)$ is non-decreasing with i.

3. Stopping condition. The algorithm stops when reaching a point that the objective function does not increase along any direction, which is a saddle point and is considered as a local optimum. When the objective function is convex, the algorithm converges to the global optima (Hunter and Lange 2000). ■

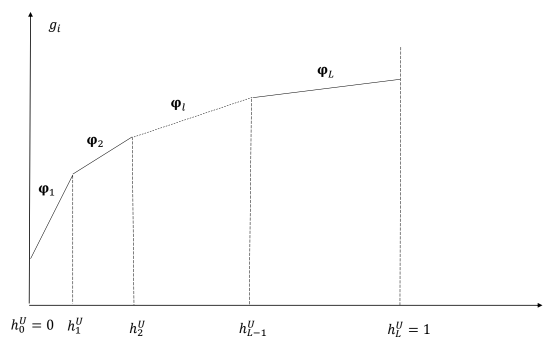


Figure A.4 Illustration of a piece-wise linear concave function.

Proof of Proposition 5

Proof. According to Theorem 1, $\rho_s^* \in (0,1)$, if and only if $W^* = \frac{\alpha \sum_i D_i - \varphi_s}{2\alpha}$; $\rho_s^* = 1$, if and only if $W^* \le \frac{\alpha \sum_i D_i - \varphi_s}{2\alpha}$.

First, we show that if $\varphi_{max} > -\alpha \sum_i D_i$, then, there exists at least one community with $\rho_i < 1$. If the optimal solution let all $\rho_i^* = 1$, then, community N could serve as the threshold community and $W^* = \sum_i D_i$. It contradicts with Theorem 1, based on which, $W^* \le \frac{\alpha \sum_i D_i - \varphi_N}{2\alpha} < \frac{\alpha \sum_i D_i + \alpha \sum_i D_i}{2\alpha} = \sum_i D_i$.

Second, we show that if $\varphi_{max} \le -\alpha \sum_i D_i$, then all communities will have $\rho_i = 1$. If not, the number of patients going to the hospital should be less than the total demand $W^* < \sum_i D_i$.

Table A.6 Hospital Revenue, Patient Surplus, and Social Welfare without and with Virtual Services.

	No VC Visits	With VC Visits
Special Price Setting	—	$\eta_\mathbf{d}\mathbf{C_{2,i}} = -\mathbf{C_2} - \boldsymbol{\Delta}\mathcal{R} + \mathbf{f_i} + \alpha\sum_\mathbf{i}\rho_\mathbf{j}\mathbf{D_j}$
Hospital	$R^0 = \sum_i D_i(-\gamma)$	$R = \sum_i D_i(-\gamma) + \sum_i D_i(1-\rho_i)(\varphi_i + \alpha\sum_j \rho_j D_j)$ $\Delta R = \sum_i D_i(1-\rho_i)(\varphi_i + \alpha\sum_j \rho_j D_j)$
Patient	$U_i^0 = \sum_i D_i(\mathcal{R}_1 - C_1 - f_i - \alpha\sum_j D_j)$	$U_i = \sum_i D_i(\mathcal{R}_1 - C_1 - f_i - \alpha\sum_j \rho_j D_j)$ $\Delta U_i = \sum_i D_i(\alpha\sum_j(1-\rho_j)D_j)$
Welfare	$S^0 = \sum_i D_i(\mathcal{R}_1 - \gamma - f_i - \alpha\sum_j D_j)$	$S = \sum_i D_i(C_1 - \gamma) + \sum_i D_i(1-\rho_i)(\varphi_i + \alpha\sum_j \rho_j D_j) + \sum_i D_i(\mathcal{R}_1 - C_1 - f_i - \alpha\sum_j \rho_j D_j)$ $\Delta S = \sum_i (1-\rho_i)D_i[\varphi_i + \alpha\sum_j(1+\rho_j)D_j]$
General Price Setting	—	$\eta_\mathbf{d}\mathbf{C_{2,i}} \le -\mathbf{C_2} - \boldsymbol{\Delta}\mathcal{R} + \mathbf{f_i} + \alpha\sum_\mathbf{j}\rho_\mathbf{j}\mathbf{D_j}$
Hospital	$R^0 = \sum_i D_i(-\gamma)$	$R = \sum_i \rho_i D_i(-\gamma) + \sum_i D_i(1-\rho_i)(\eta_d C_{2,i} + C_2 - \beta_i)$ $\Delta R = \sum_i D_i(1-\rho_i)(\eta_d C_{2,i} + C_2 - \beta_i - (-\gamma))$
Patient	$U^0 = \sum_i D_i(\mathcal{R}_1 - C_1 - f_i - \alpha\sum_j D_j)$	$U = \sum_i (1-\rho_i)D_i(\mathcal{R}_2 - C_1 - \eta_d C_{2,i} - C_2) + \sum_i \rho_i D_i(\mathcal{R}_1 - C_1 - f_i - \alpha\sum_j \rho_j D_j)$ $\Delta U = \sum_i (1-\rho_i)D_i(-\eta_d C_{2,i} - C_2 - \Delta\mathcal{R} + f_i + \alpha\sum_j(1+\rho_j)D_j)$
Welfare	$S^0 = \sum_i D_i(\mathcal{R}_1 - \gamma - f_i - \alpha\sum_j D_j)$	$S = \sum_i \rho_i D_i(\mathcal{R}_1 - \gamma - f_i - \alpha\sum_j \rho_j D_j) + \sum_i (1-\rho_i)D_i(\mathcal{R}_2 - \beta_i)$ $\Delta S = \sum_i (1-\rho_i)D_i[\varphi_i + \alpha\sum_j(1+\rho_j)D_j]$

However, $\forall\ j \in I$, since $\varphi_j \le -\alpha\sum_i D_i$, then, $\frac{\alpha\sum_i D_i - \varphi_j}{2\alpha} \ge \sum_i D_i$, and we have $W^* < \frac{\alpha\sum_i D_i - \varphi_s}{2\alpha}$. Based on Theorem 1, it suggests that $\rho_j = 1$. Contradict.

Therefore, when $\varphi_{max} > -\alpha\sum_i D_i$, there exist patients who receive care at home. When there are patients staying at home, we can see from Table A.6 that all of the three parties are strictly better off.

Next, we show that the hospital earns strictly more money from every at-home patient.

Patients' IC constraints require $\eta_d C_{2,i} + C_2 \le -\Delta\mathcal{R} + f_i + \alpha\sum_j \rho_j D_j$, $\forall i \in I$. To obtain the maximum revenue, $\eta_d C_{2,i} + C_2 = -\Delta\mathcal{R} + f_i + \alpha W^*$. For each patient staying at their own community, the hospital now gains $\eta_d C_{2,i} + C_2 - \beta_i$. First, we show that the threshold community K have $\alpha\sum D_i + \varphi_K > 0$ because otherwise, $\alpha\sum D_i + \varphi_K \le 0$ leads to $W^* = \frac{\alpha\sum_i D_i - \varphi_K}{2} \ge \frac{\alpha\sum_i D_i + \alpha\sum_i D_i}{2} = \sum D_i$, which contradicts with the condition that there exist at-home patients. Second, since $\varphi_i \ge \varphi_K$ for each community i that is optimal to stay at home, we conclude for each at-home community $\varphi_i > -\alpha\sum_i D_i$. Compared to the previous gain $-\gamma$ from community i, the hospital now gains

$\eta_d C_{2,i} + C_2 - \beta_i$ which is strictly greater than $-\gamma$, and it collects $\eta_d C_{2,i} + C_2 - \beta_i - (-\gamma) = \varphi_i + \alpha W^*$ more money from every at-home patient. $\varphi_i + \alpha W^* \geq \varphi_K + \alpha W^*$, and the latter is positive because $\alpha W = (\alpha \sum_i D_i - \varphi_K)/2$ and $\varphi_K + \alpha \sum_i D_i > 0$. ■

Similar to the profit-maximizing problem, the social-welfare maximization problem of consideration is equivalent to the following model, and all the results are derived from it. The proof follows that in Proposition A.3.

$$\begin{aligned} \max_{\rho_i, C_{2,i}} \quad & \sum_{i \in I} (1-\rho_i) D_i [\varphi_i + \alpha \sum_{j \in I} (1+\rho_j) D_j], \\ \text{s.t.} \quad & \text{IC constraints (1)} - (3). \end{aligned}$$

Proof of Proposition 7

Proof. For a linear nurse coordination cost function, we obtain:

$$L_{\text{Welfare}} = \sum_i (1-\rho_i) D_i [\varphi_i + \alpha \sum_j (1+\rho_j) D_j] + \lambda_i \rho_i D_i + \mu_i (1-\rho_i) D_i.$$

The FOC suggests that

$$\frac{\partial L_{\text{Welfare}}}{\partial \rho_i} = -D_i(\varphi_i + \alpha \sum_j \rho_j D_j) + \alpha D_i \sum_j D_j (1-\rho_j) - \alpha D_i \sum_j D_j + \lambda_i D_i - \mu_i D_i = 0.$$

As as result, $\sum_j \rho_j D_j = \frac{-\varphi_i + \lambda_i - \mu_i}{2\alpha}$. Note that $W^{\text{Wel}} = \sum_j \rho_j D_j$, similar to the revenue-maximizing case, we have

$$\begin{aligned} \rho_i^{\text{Wel}} \in (0,1), \quad & \text{if and only if} \quad W^{\text{Wel}} = \frac{-\varphi_i}{2\alpha}; \\ \rho_i^{\text{Wel}} = 1, \quad & \text{if and only if} \quad W^{\text{Wel}} \leq \frac{-\varphi_i}{2\alpha}; \\ \rho_i^{\text{Wel}} = 0, \quad & \text{if and only if} \quad W^{\text{Wel}} \geq \frac{-\varphi_i}{2\alpha}. \end{aligned}$$

Therefore, a threshold structure also applies to the social welfare optimization problem with the threshold K^{Wel}.

Next, we show $W^{\text{Wel}} < W^{\text{Rev}}$ by considering the following cases:

- Case 1: $K^{\text{Rev}} > K^{\text{Wel}}$.

As $W = \rho_K D_K + \sum_{j<K} D_j$, $K^{\text{Rev}} > K^{\text{Wel}}$ leads to $W^{\text{Rev}} > W^{\text{Wel}}$.

- Case 2: $K^{\text{Rev}} = K^{\text{Wel}}$.

$$W^{\text{Rev}} = \frac{\alpha \sum_i D_i - \varphi_{K^{\text{Rev}}}}{2\alpha} > \frac{\alpha - \varphi_{K^{\text{Rev}}}}{2\alpha} = \frac{\alpha - \varphi_{K^{\text{Wel}}}}{2\alpha} = W^{\text{Wel}}.$$

- Case 3: $K^{\text{Rev}} < K^{\text{Wel}}$.

As $W = \rho_K D_K + \sum_{j<K} D_j$, $K^{\text{Rev}} < K^{\text{Wel}}$ leads to $W^{\text{Rev}} < W^{\text{Wel}}$. Since $\varphi_{K^{\text{Rev}}} < \varphi_{K^{\text{Wel}}}$,

$$W^{\text{Rev}} = \frac{\alpha \sum_i D_i - \varphi_{K^{\text{Rev}}}}{2\alpha} > \frac{\alpha - \varphi_{K^{\text{Rev}}}}{2\alpha} > \frac{\alpha - \varphi_{K^{\text{Wel}}}}{2\alpha} = W^{\text{Wel}},$$

which contradicts with $W^{\text{Rev}} < W^{\text{Wel}}$.

Introduce the superscript "Rev" to refer to the revenue maximization scenario. $\Delta_{\text{Patients}} = \alpha H \sum_j D_j$. Since $H^{\text{Wel}} > H^{\text{Rev}}$, then, $\Delta^{\text{Wel}}_{\text{Patients}} > \Delta^{\text{Rev}}_{\text{Patients}}$. Moreover, $\Delta^{\text{Wel}}_{\text{Hospital}} < \Delta^{\text{Rev}}_{\text{Hospital}}$ and $\Delta^{\text{Wel}}_{\text{Welfare}} > \Delta^{\text{Rev}}_{\text{Welfare}}$ follow from the fact that ρ^{Rev} maximizes revenue (Δ_{Hospital}) and ρ^{Wel} maximizes social welfare (Δ_{Welfare}). ∎

Proof of Proposition 6

Proof. Let Δ_{Hospital} represent the difference of revenues with and without offering virtual services, and Δ_{Patients} be the change of patient surplus. We introduce superscript "Wel" to refer to the results obtained from social welfare maximization. According to Proposition 7, $\rho_i^{\text{Wel}} \in (0,1)$, if and only if $W^{\text{Wel}} = \frac{-\varphi_i}{2\alpha}$; $\rho_i^{\text{Wel}} = 1$, if and only if $W^{\text{Wel}} \le \frac{-\varphi_i}{2\alpha}$.

First, we show that if $\varphi_{max} > -2\alpha \sum_i D_i$, then there exists at least one community with $\rho_i < 1$. If $\rho_i = 1$ for all $i \in I$, community N could serve as the threshold community and $W^{\text{Wel}} = \sum_i D_i$, which conflicts with $W^{\text{Wel}} \le \frac{-\varphi_{max}}{2\alpha} < \frac{2\alpha \sum_i D_i}{2\alpha} = \sum_i D_i$.

Second, we show that if $\varphi_{max} \le -2\alpha \sum_i D_i$, then, all communities have $\rho_i = 1$. If not, then, the number of patients going to the hospital should be less than the total demand rate $W^{\text{Wel}} < \sum_i D_i$. Now, $\forall\ j \in I$, since $\varphi_j \le -\alpha \sum_i D_i$, then, $\frac{-\varphi_j}{2\alpha} \ge \sum_i D_i$, we have $W^{\text{Wel}} \le \frac{-\varphi_s}{2\alpha}$, and according to Proposition 7, $\rho_j = 1$. Contradict.

For the change of social welfare without and with VC visits, defined as $\Delta_{\text{Welfare}} = \sum_i D_i(1 - \rho_i)(\varphi_i + \alpha \sum_i D_i + \alpha W^{\text{Wel}}) = \sum_i D_i(1-\rho_i)\varphi_i + \alpha H^{\text{Wel}}(\sum_i D_i + W^{\text{Wel}})$. If $H^{\text{Wel}} > 0$, $\Delta_{\text{Welfare}} > 0$. One the other hand, $\Delta_{\text{Welfare}} = 0$ leads to $H^{\text{Wel}} = 0$. Thus, the change in social welfare is strictly positive when there exists at least one patient staying at home.

∎

Proof of Corollary 1

Proof. For each community i with $\varphi_i \ge -\alpha W^{\text{Wel}}$, the medical institution is willing to provide them VC visits. Note here W^{Wel} is invariant with $C_{2,i}$. When $\eta_d C^{\text{Wel}}_{2,i} \le -C_2 - \Delta\mathcal{R} + f_i + \alpha W^{\text{Wel}}$, patients are optimal to choose the virtual service option. When $-C_2 - \Delta\mathcal{R} - \gamma + \beta_i \le \eta_d C_{2,i}^{\text{Wel}}$, the medical institution would like to offer virtual services. In summary, any price $\eta_d C_{2,i}^{\text{Wel}}$ between $-C_2 - \Delta\mathcal{R} - \gamma + \beta_i$ and $-C_2 - \Delta\mathcal{R} + f_i + \alpha W^{\text{Wel}}$ would be a feasible price. If $\varphi_i < -\alpha W^{\text{Wel}}$, the hospital is forced to offer VC visits and the equilibrium pricing is $\eta_d C_{2,i}^{\text{Wel}} = -C_2 - \Delta\mathcal{R} + f_i + \alpha W^{\text{Wel}}$, which follows patients' IC constraints.

Under the condition that $\varphi_{max} > -2\alpha \sum_i D_i$:

1. By setting $\eta_d C_{2,i} = -C_2 - \Delta\mathcal{R} + f_i + \alpha W^{\text{Wel}}$, the hospital is strictly better off. For the hospital, since $\Delta_{\text{Hospital}} = \sum_i D_i(1-\rho_i)\varphi_i + \alpha H^{\text{Wel}} W^{\text{Wel}}$. $H^{\text{Wel}} > 0$ leads to $\Delta_{\text{Hospital}} > 0$. Otherwise, $H^{\text{Wel}} = 0$ leads to $\rho_i = 1$, $\forall i$ and thus $\Delta_{\text{Hospital}} = 0$. Therefore, the hospital is strictly better off if there exists at least one patient staying at home. Social welfare, as the sum of the two, is also positive.

However, whether the hospital receives more money from each community is not guaranteed under the social maximizer. In fact, the hospital receives more money when $-\alpha W^{\text{Wel}} \le \varphi_i$, because $-C_2 - \Delta\mathcal{R} - \gamma + \beta_i \le \eta_d C_{2,i} = -C_2 - \Delta\mathcal{R} + f_i + \alpha W^{\text{Wel}}$. However, for communities with $\varphi_i < -\alpha W^{\text{Wel}}$, the hospital gains less. Such communities exist when $W^{\text{Wel}} > 0$, i.e., $\exists \quad \rho_i^{\text{Wel}} \in (0,1]$, which means $\varphi_i \le -2\alpha W^{\text{Wel}} < -\alpha W^{\text{Wel}}$.

2. For the communities with $\varphi_i \in [-2\alpha W^{\text{Wel}}, -\alpha W^{\text{Wel}})$, $\eta_d C_{2,i} = -C_2 - \Delta\mathcal{R} + f_i + \alpha W^{\text{Wel}} < -C_2 - \Delta\mathcal{R} - \gamma + \beta_i$ and $0 \le \rho_i < 1$, the hospital is losing money serving them. For the communities with $\varphi_i \in [-\alpha W^{\text{Wel}}, +\infty)$, when we assign $\eta_d C_{2,i} = -C_2 - \Delta\mathcal{R} - \gamma + \beta_i$, which is a feasible price, the hospital does not gain more from serving them. Overall, the hospital's gain is negative.

■

Corollary A.2. *When priced at $\eta_d C_{2,i}^{Wel} = -C_2 - \Delta\mathcal{R} + f_i + \alpha W^{Wel}$, if $\forall i$, $\varphi_i \ge 0$, then, $\rho_i^{Wel} = 0$, i.e., all patients should stay at home under social welfare maximization.*

For all patients to stay at home, compared to the condition $\varphi_j \ge \alpha \sum_{i\in I} D_i$ in Corollary A.1 in the revenue maximization case, the condition of the social planner (Corollary A.2) is relaxed — as long as all of their marginal utility changes are nonnegative.

Proof of Theorem 4

Proof. For part (a), patients from community j being optimal to stay at home means $\eta_f C_{2,f} + C_2 < f_j + \alpha W - \Delta\mathcal{R}$. Assume not, patients from community i are also optimal to go to hospital, i.e., $f_i + \alpha W - \Delta\mathcal{R} \le \eta_f C_{2,f} + C_2$. These results lead to $f_i < f_j$ for any $\eta_f C_{2,f} + C_2$ and W, which contradicts with $f_i > f_j$. Part (b) naturally follows and the threshold distance is $F = \eta_f C_{2,f} + C_2 - \alpha W + \Delta\mathcal{R}$.

■

PROPOSITION A.4. *When $\eta_d = 0$, and $\eta_f = 0$, the medical institution is not willing to offer VC visits unless C_2 satisfies the condition described by inequality (A.8).*

Proof. When C_2 is a constant, firstly, when F increases, W also increases, so there exists a unique solution for $F + \alpha W = C_2 + \Delta\mathcal{R}$, and then, the equilibrium patient strategy can be obtained. F and W can be found by applying the monotone property in Theorem 4 through the following procedure: Test $F \in (f_{i-1}, f_i]$, for $i = N+1, N, .., 1$, where $f_{N+1} = \infty, f_0 = 0$, and stop when feasible F and W are found.

- Step 1: Set initial $W = \sum_i D_i$. Start with $i = N+1$, if $C_2 > f_i + \alpha W - \Delta\mathcal{R}$ (denoted as "ineq*"), i.e., $F > f_N$, which means feasible F and W have been found, then, all patients are optimal to go to the hospital. Else, go to step 2.
- Step 2: (to determine F) Consider community i, if $C_2 > f_{i-1} + \alpha(W - D_i) - \Delta\mathcal{R}$, stop and $F \in (f_{i-1}, f_i]$, go to step 3. Else, continue with $i = i-1$ and $W = W - D_i$.
- Step 3: (to determine ρ_K) Now, denote i as K, a threshold community K. A mixed strategy only exists when $F = f_K$. Since $f_K = C_2 - \alpha(W - D_K + \rho_K D_K) + \Delta\mathcal{R}$, we obtain $\rho_K = \max\{0, (-f_K + C_2 + \Delta\mathcal{R} - \alpha(W - D_K))/D_K\}$. Notice that $\rho_K < 1$ because of (ineq*). If $\rho_K > 0$, $F = f_K$, and if $\rho_K = 0$, $F \in (f_{K-1}, f_K)$.

Based on individual patients' choices, the revenue becomes:

$$
\begin{aligned}
obj = C_2 \left(\sum_{j: f_j > f_K} D_j + (1-\rho_K) D_K \right) - \sum_{j: f_j > f_K} g_j(D_j) - g_K((1-\rho_K)D_K) \\
- \gamma[\sum_{j: f_j < f_K} D_j + \rho_K D_K - W_e]^+ + \gamma[\sum_{j \in I} D_j - W_e]^+.
\end{aligned} \tag{A.7}
$$

After finding out the patient strategy and F, then, the medical institution is willing to offer VC visits when

$$
C_2 \geq \frac{\sum_{j: f_j > F} g_j(D_j) + \sum_{j: f_j = F} g_j((1-\rho_j)D_j) + \gamma[\sum_{j: f_j < F} D_j + \rho_K D_K - W_e]^+ - \gamma[\sum_{j \in I} D_j - W_e]^+}{\sum_{j: f_j > F} D_j + \sum_{j: f_j = F} (1-\rho_j) D_j}. \tag{A.8}
$$

■

LEMMA A.4. *When $\eta_d = 0$ and $\eta_f \neq 0$, the medical institution is able to find equilibrium patient strategies and the optimal $C_{2,f}$ in $O(N)$ time when $g(\cdot)$ is finite piece-wise smooth.*

Proof. For mathematical simplicity, we call K the threshold community when $F \in (f_{K-1}, f_K]$. If $\rho_K \in (0,1)$, $F = f_K$ and $\eta_f C_{2,f} + C_2 = f_i + \alpha W - \Delta\mathcal{R}$. If $\rho_K = 0$, $F \in (f_{K-1}, f_K)$. To maximize revenue, the medical institution would set $F = f_K - \epsilon$, and $\eta_f C_{2,f} + C_2 = f_K - \epsilon + \alpha W - \Delta\mathcal{R}$, for a small amount ϵ. Therefore, the equilibrium price is $\eta_f C_{2,f} + C_2 = f_K + \alpha W - \Delta\mathcal{R}$. Notice that if $\rho_K = 1$, K is not the threshold community. We first find the threshold community and the optimal strategy ρ_K:

$$\begin{aligned}
K &= \mathrm{argmax}_{i=0,1,..,N}\, obj(i), \\
obj(i) &= \max_{\rho_i}(\eta_f C_{2,f} + C_2)\left(\sum_{j:f_j>f_i} D_j + (1-\rho_i)D_i\right) - \sum_{j:f_j>f_i} g_j(D_j) - g_i((1-\rho_i)D_i) \\
&\quad - \gamma[\sum_{j:f_j<f_i} D_j + \rho_i D_i - W_e]^+ + \gamma[\sum_{j\in I} D_j - W_e]^+,
\end{aligned} \tag{A.9}$$

where $W = \sum_{j:f_j<f_i} D_j + \rho_i D_i$ and $\eta_f C_{2,f} + C_2 = f_i + \alpha W - \Delta\mathcal{R}$.

The first term is a quadratic function. Thus, this unconstrained optimization problem with one decision variable ρ_i is easy to solve. For the linear case where $g_i((1-\rho_i)D_i) = \beta_i(1-\rho_i)D_i$, and $W_e = 0$, we have a close-form optimal solution

$$\rho_i = \max\{0, \min\{1, \frac{\alpha(H_i^C + D_i - W_i^C) - \varphi_i}{2\alpha D_i}\}\},$$

where $\varphi_i = f_i - \beta_i + \gamma - \Delta\mathcal{R}$, $W_i^C = \sum_{j:f_j<f_i} D_j$ and $H_i^C = \sum_{j:f_j>f_i} D_j$.

When $g(\cdot)$ is smooth or finite piece-wise smooth, we are able to obtain the optimal ρ_i by using the first derivative or iterating all discontinuous points. ■

Proof of Proposition 8

Proof. The medical institution would like to offer the service when there exists i such that $obj(i) > 0$. This is equivalent to there exists i such that $\rho_i < 1$, for

$$\rho_i = \max\{0, \min\{1, \frac{\alpha(H_i^C + D_i - W_i^C) - \varphi_i}{2\alpha D_i}\}\},$$

which means $\varphi_i > \alpha(\sum_{f_j>f_i} D_j - \sum_{f_j\leq f_i} D_j)$.

The reason for the equivalence is as follows. Assume the optimal solution of ρ_k when k is the threshold community is ρ_k^*. If the two statements are not equivalent, then when all $\rho_k^*=1,\ \forall k\in I$, there exists $obj(i)>0$. Observe that

$$\begin{aligned} obj(i,\rho_i^*=1) &< obj(i+1,\rho_{i+1}=0) \quad \text{(Definition of } obj(i,\rho_i) \text{ in (A.9))}\\ &< obj(i+1,\rho_{i+1}=1) \quad \text{(Assumption } \rho_k^*=1,\ \forall k\in I)\\ &< \ldots < obj(N,\rho_N=1)=0. \end{aligned}$$

Contradict.

Under the condition that $\exists\ \varphi_i>\alpha(\sum_{f_j>f_i}D_j-\sum_{f_j\le f_i}D_j)$, there is at least one patient optimal to stay at home, and the congestion level at the medical institution decreases comparing to without VC visits. Hence, the go-to-hospital option has an increased utility. Subsequently, the patients who are optimal to go to the hospital are strictly better off. Next, consider the patients who prefer to stay at home. Given that they obtain larger utilities from the at-home option than that from the go-to-hospital option, they are also strictly better off. ■

Proof of Proposition 9

Proof. Social maximizer is to find:

$$\begin{aligned} \max_{C_2}\ & \sum_i D_i(1-\rho_i)(-\Delta\mathcal{R}+f_i+\alpha\sum_j D_j)-\sum_i g_i((1-\rho_i)D_i)-\gamma[\sum_i\rho_i D_i-W_e]^+ +\gamma[\sum_{j\in I}D_j-W_e]^+\\ & +\sum_i\rho_i D_i(\alpha\sum_j(1-\rho_j)D_j),\\ \text{s.t.}\ & \text{IC constraints (1)}-\text{(3)}. \end{aligned}$$

Follow the same logic with revenue maximizer, the social welfare gain from the new service is $obj^{\text{Wel}}(K)$:

$$\begin{aligned} K &= \text{argmax}_{i=0,1,..,N}\, obj^{\text{Wel}}(i)\\ obj^{\text{Wel}}(i) &= \max_{\rho_i}\Big(\sum_{j:f_j>f_i}D_j(-\Delta\mathcal{R}+f_j+\alpha\sum_j D_j)+(1-\rho_i)D_i(-\Delta\mathcal{R}+f_i+\alpha\sum_j D_j)\\ &\quad+\sum_{j:f_j<f_i}D_j(\alpha\sum_j D_j-\alpha W)+\rho_i D_i(\alpha\sum_j D_j-\alpha W)\\ &\quad-\sum_{j:f_j>f_i}g_j(D_j)-g_i((1-\rho_i)D_i)-\gamma[\sum_{j:f_j<f_i}D_j+\rho_i D_i-W_e]^+ +\gamma[\sum_{j\in I}D_j-W_e]^+\Big). \end{aligned}$$

$$= \max_{\rho_i} \left(\sum_{j:f_j>f_i} (\varphi_j + \sum_{j\in I} D_j) D_j + (1-\rho_i)(\varphi_i + \sum_{j\in I} D_j) D_i \right) + \alpha W (\sum_{j\in I} D_j - W) \quad \text{(all linear case)}$$

where $W = \sum_{j:f_j<f_i} D_j + \rho_i D_i$, and $\varphi_j = f_j - \beta_j + \gamma - \Delta\mathcal{R}$.

The close-form optimal solution is

$$\rho_i = \max\{0, \min\{1, \frac{-2\alpha W_i^C - \varphi_i}{2\alpha D_i}\}\},$$

where $\varphi_i = f_i - \beta_i + \gamma - \Delta\mathcal{R}$, and $W_i^C = \sum_{j:f_j<f_i} D_j$.

Following the same logic as that in the proof of Proposition 8, social welfare begins to increase when $\frac{-2\alpha W_i^C - \varphi_i}{2\alpha D_i} < 1$, i.e., when there exists a community i with $\varphi_i > -2\alpha \sum_{f_j \le f_i} D_j$. ■

Appendix F Additional Numerical Studies

We analyze the number of registered cancer patients by city and county (zip code indexed) throughout calender year 2018 to identify the service region I covered by UFHCC and to estimate the potential patient demands D_i across regions. We crafted one numerical example based on the real-world data presented in Table A.7. Additional model parameters are displayed in Table A.8. We introduce $\beta_i = \text{ratio}_\beta * \text{distance}_i$, that is, the cost of dispatching nurses is linear to the distance between the patient home and the community clinic, and similarly, we define a travel burden-to-distance ratio, suggesting that the travel cost of patients is linear to the distance between patient homes and the medical center

F.1 Sensitivity Analysis

Here we present a sensitivity analysis to explore how the decisions vary based on the parameter setting. First, we consider the demand variation, which is captured by the demand-to-population ratio. When it decreases, i.e., not all registered patients do follow-ups, the threshold community is still Marion and 41.5% of patients therein go to the medical center, and the rest stay at home (51.0% market size). Without the capacity cap, when the demand-to-population ratio increases from 1 to 2, 44.8% of Marion patients, and 50.3% of total patients are optimal to stay at home. It suggests that the relative market size of telehealth is decreasing with the increase in demand rate. This seemingly counter-intuitive observation can be explained by the high congestion cost perceived by patients, and the medical institution can take advantage of this to create congestion that raises patients' willingness to pay for telehealth. The negative externality of congestion perceived by the

Table A.7 Sample data and optimal solutions.

County Name	D_i	f_i	$f_i - \beta_i$	Rev. Max. ρ^* (Base case:	Social Max. ρ^* $\theta = 1,000$)	Rev. Max.ρ^* (Large θ:	Social Max. ρ^* $\theta = 10,000$)
Alachua, FL	292	2.0	0.0	1	.	1	.
Levy, FL	83	22.4	0.0	1	.	1	.
Clay, FL	265	25.1	0.0	1	.	1	.
Bradford, FL	41	28.3	0.0	1	.	1	.
Union, FL	18	29.1	0.0	1	.	1	1
Gilchrist, FL	29	31	0.0	1	.	1	1
Putnam, FL	135	35	0.0	1	.	1	.
Dixie, FL	32	52	0.0	1	.	1	1
Citrus, FL	424	58.8	19.4	1	.	1	.
Baker, FL	31	59	26.2	1	.	1	1
Lafayette, FL	11	67.8	31.1	1	.	1	1
Marion, FL	828	38.3	34.7	.32	.	.	.
Suwannee, FL	77	68	41.7	1	.	1	.
Hamilton, FL	21	75.3	43.0	1	.	1	1
Sumter, FL	583	73.4	45.2	.	.	.	.
Columbia, FL	104	46.8	46.0	.	.	1	.
Madison, FL	32	105	72.1	.	.	1	.
Taylor, FL	36	100	98.4	.	.	1	.
Jefferson, FL	27	129	101.8	.	.	1	.
Wakulla, FL	40	143	115.6	.	.	1	.
Leon, FL	305	144	140.3	.	.	.	.
Gadsden, FL	67	176	140.7	.	.	1	.

Table A.8 Summary of the model parameter setting.

Parameter	Baseline	Range	Unit
α	1	$[1, 10)$	\$/pp
γ	1	$[1, 10)$	\$
ratio$_\beta$	1	$[1, 1,000]$	\$/mile
demand-to-discharge ratio	1	$[0.5, 2]$	–
travel burden-to-distance ratio	1	$[1, 10]$	\$/mile
W_e	0	$[0, 10,000]$	–

medical institution and patients have different effects. When the medical institution is sensitive to hospital congestion, for example, γ increases from 1 to 10, 43.1% of Marion patients should go to the medical center. Whereas when patients are more sensitive to congestion, for example, α increases from 1 to 10, 45.6% of Marion patients should go to the medical center. Note that Marion patients always adopt mixed strategies, which agrees with our previous observation that a large demand size is more likely to induce mixed strategy optimal solutions.

A smaller nurse coordination cost (for example, a per-mile nurse dispatch cost ratio$_\beta = 0.5$ in contrast to 1) would cause more patients to stay at home. Three counties become more beneficial than Marion (less beneficial in the base case), and they become optimal to stay at home. Meanwhile, only 0.2% of Marion patients are still optimal to stay at home, and overall half of the patients are optimal at home receiving VC services. A larger nurse coordination cost (for example, $\text{ratio}_\beta = 2$) also leads to changes in marginal benefits and ranking of communities. In this case, it indeed causes more patients to go to the medical center, and 45.5% of total patients are optimal to stay at home. On the other hand, when the unit distance travel burden increases from 1 to 10, i.e., the transportation cost of patients f_i becomes higher, 55.0% of all patients stay at home.

F.2 Concave Cost Function

In addition, we consider a concave cost function with a non-zero set-up cost, following a previously published work on profit-maximizing cost allocation problem for firms using cost-based pricing (Pavia 1995). In this numerical study, we use a set-up cost $\theta = 1,000$ in the base case. A piece-wise linear concave function as in (Hu et al. 2019) with $L = 3$, $\text{coef}_1 = 1$ and coef_l being the reverse order statistics of L uniform random numbers in $[0.6, 1]$, for all $1 < l \leq L$. This yields $g(D) = 1000 + \beta_i \min\{D, 100\} + 0.7\beta_i[\min\{D, 200\} - 100]^+ + 0.6\beta_i[D - 200]^+$, as shown in Figure A.5. Consequently, for each community, the gradients of the cost function are the coefficient of segments times the marginal cost β_i used in the linear setting, and all parameters stay the same as the linear baseline in Table A.8. We provide optimal solutions of the base case with $W_e = 1,000$, $\theta = 1,000$, (see Figure A.6) and another case with a large θ: $\theta = 10,000$ for comparison in Table A.7.

The impact of being a strictly concave rather than linear cost function on the optimal solution is also minimal. This is because only communities employing mixed strategies have nurse coordination costs differing from the ones under the linear approximations. Since the concave cost function is computationally expensive, one can initialize the heuristic algorithm (Algorithm 2) with the optimal solution obtained from the linear approximation to gain efficiency.

In contrast, the set-up cost poses a significant impact on the optimal solution. When set-up costs are high, for example, $\theta = 10,000$, we observe that the solution structure with a threshold community using a mixed strategy no longer exists. Instead, all communities adopt pure strategies, and more communities are optimal to go to the hospital because of the high set-up cost.

Whereas a single factor, marginal gain φ_i, is sufficient to decide the threshold community in the linear case, when a general cost function is assumed, both marginal gain φ_i and demand size D_i

should be taken into consideration. Intuitively, with similar marginal gains, the medical institution would prefer communities with a larger demand to stay at home. Generally, the communities with good profitability of telehealth (the product of marginal profit and demand is high) are optimal to stay at home. However, the marginal gain still plays a leading role — if the marginal profit is too small, no matter how large the population is, the community is still optimal to go to the hospital. In contrast, when it comes to the social welfare optimization, D_i is an important factor — a social planner would only let a community as a whole to go to the hospital when its population size is small.

Lastly, our conclusions that (1) a social planner would like more patients to stay at home, and (2) under the given conditions in Proposition 7, $\Delta^{\text{Wel}}_{\text{Patients}} > \Delta^{\text{Rev}}_{\text{Patients}}$, $\Delta^{\text{Wel}}_{\text{Hospital}} < \Delta^{\text{Rev}}_{\text{Hospital}}$, and $\Delta^{\text{Wel}}_{\text{Welfare}} > \Delta^{\text{Rev}}_{\text{Welfare}}$ remain valid for the general concave nurse coordination cost functions. The difference of the optimal solution comparing a concave cost function and its linear approximation is small, and the readers are referred to the Appendix for more numerical results regarding the benefits brought by VC visits for hospital revenue, patient surplus, and social welfare under different settings.

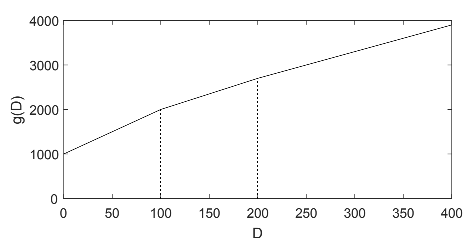


Figure A.5 Sample piece-wise linear concave function

F.3 Non-Zero Congestion Penalty

We test different non-zero congestion penalty threshold values W_e. To our surprise, this non-zero threshold is negligible when the medical institution's sensitivity to congestion is small (such as $\gamma = 1$ in the baseline setting). When the medical institution's sensitivity to congestion becomes larger, for example, $\gamma = 10$, this threshold comes into play and affects the optimal solutions — a zero congestion penalty threshold ($W_e = 0$) and a non-zero congestion penalty threshold $W_e = 1000$ favor less patients coming to the central facility compared to $W_e = 2000$, where the whole population generates a demand $D_{\text{All}} = 3481$ in the baseline setting.

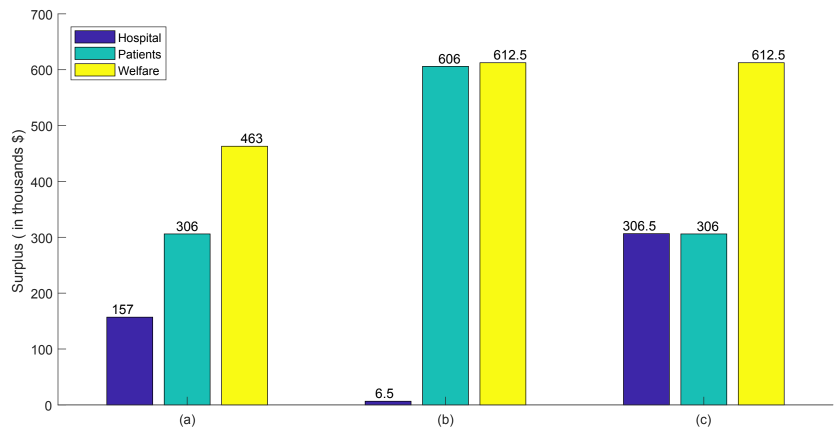


Figure A.6 Results for piece-wise linear concave cost functions (Base case).Benefits brought by telehealth for hospital revenue, patient surplus, and social welfare under different settings: (a) revenue maximization, (b) social welfare maximization, and (c) social welfare maximization with increased in-person care price C_1

For the sensitivity analysis regarding $W_e \neq 0$, we provide the following discussion. With the non-linear congestion cost $-\gamma[W - W_e]^+$, denote $R_e(\cdot)$, the revenue-maximizing objective function as a function of congestion W, given a non-zero $W_e = w_e < \sum_i D_i$. In addition, let R^* be the optimal objective function, given $W_e = w_e$. This is because the change of the objective function does not affect patients' incentive compatibility, and thus, once the prices are given, W is given, and ρ_i's can be uniquely determined (see the proof of Proposition A.3).

(a) When $W_e = 0$ (reduced to a linear term $-\gamma W$ with $\gamma \neq 0$), we denote the optimal congestion as W_0^*. Since W_0^* is always a feasible solution, $R^* \geq R_e(W_0^*)$. Moreover, denote the objective function when $W_e = 0$ as $R_0(\cdot)$, then we have $R_e(W_0^*) \geq R_0(W_0^*)$. Thus, $R_0(W_0^*)$ serves as a lower bound of R^*.

(b) When $W_e = \sum_i D_i$ (equivalent to $\gamma = 0$), we denote the optimal congestion as W_∞^*, and the objective function $R_\infty(\cdot)$. We know $W_0^* \leq W_\infty^*$ and $R_e(W_0^*) \leq R_e(W_\infty^*)$. Because we drop a non-positive term in $R_\infty(\cdot)$, thus, for any W, we have $R_e(W) \leq R_\infty(W) \leq R_\infty(W_\infty^*)$. We obtain an upper bound of R^*.

Therefore, the optimal objection function R^* is between $R_0(W_0^*)$ and $R_\infty^*(W_\infty^*)$. For the revenue-maximizing problem, if the impact of γ (the sensitivity of congestion from the medical institution's side) is small, then the difference between $R_0(W_0^*)$ and $R_\infty(W_\infty^*)$ is small, i.e., the distance between the lower and upper bound of R^* is small. As a result, the impact of W_e is small. Note that the impact of γ to the hospital's revenue is in proportion to the number of patients, whereas the impact of α (the sensitivity of congestion from the patient's side) is in proportion to the square of the number of patients.