Managing capacity for virtual and office appointments in chronic care

Xiao Yu[1] · Armagan Bayram[1]

Abstract

Patients living with a chronic disease often require regular appointments and treatments. Due to the constraints on the availability of office appointments and the capacity of physicians, access to chronic care can be limited; consequently, patients may fail to receive the recommended care suggested by clinical guidelines. Virtual appointments can provide a cost-effective alternative to traditional office appointments for managing chronic conditions. Advances in information technology infrastructure, communication, and connected medical devices are enabling providers to evaluate, diagnose, and treat patients remotely. In this study, we build a capacity allocation model to study the use of virtual appointments in a chronic care setting. We consider a cohort of patients receiving chronic care and model the flow of the patients between office and virtual appointments using an open migration network. We formulate the planning of capacity needed for office and virtual appointments with a newsvendor model to maximize long-run average earnings. We consider differences in treatment and diagnosis effectiveness for office and virtual appointments. We derive optimal capacity allocation policies and implement numerical experiments. With the model developed, capacity decisions for office and virtual appointments can be made more systematically with the consideration of patient disease progressions.

Highlights

- We consider an operational problem that deals with the effective allocation of capacity among different types of appointments.
- We integrate the operational decisions (i.e., capacity allocation) with the clinical operations (i.e., treatment and diagnosis of patients, disease progression).
- We show how the expected number of patients at different appointments changes with respect to some of the model parameters.
- We find that although virtual appointments are not as effective as office appointments, they can have equal importance to office appointments due to their lower costs.
- We propose efficient and practical solutions that would help clinics in their capacity allocation decisions.

✉ Armagan Bayram
armagan@umich.edu

Xiao Yu
xiaoyuu@umich.edu

1 Industrial and Manufacturing Systems Engineering, University of Michigan-Dearborn, Dearborn, MI 48128, USA

1 Introduction

Chronic care involves the treatment and monitoring of pre-existing and long-term diseases such as diabetes, high blood pressure, asthma, Alzheimer's disease, and cardiovascular disease [10]. In the U.S., 45% of the population has at least one chronic disease, and the cost of chronic care constitutes over 75% of the entire health care spending in the U.S. [25, 48]. Given that the population is increasing and aging, the need for chronic care in the future will increase faster. Current care processes are insufficient to address this coming mismatch in supply and demand [24]. To improve patient access to chronic care and to reduce their burden, health care providers increasingly rely on virtual appointments as a new alternative way to provide effective and consistent long-term care. Virtual appointments,

consisting of e-mail, phone, and online consultations, can improve patient access and ensure continuity of care and, consequently, better outcomes [12, 42].

Virtual appointments can be used as a substitute for, or be complementary to office appointments, and they can take many different forms. For example, virtual appointments can be used for diagnosis only, for treatment only, and for both treatment and diagnosis, similar to office appointments [7]. More specifically, through virtual appointments that provide diagnosis only, chronic-care patients can be monitored remotely in real time and updates regarding the patients' status can be obtained [3, 37]. Through virtual appointments that provide treatment only, educational support and reliable resources can be provided to patients without diagnosing their status [22]. Finally, virtual appointments can also be used to provide both diagnosis and treatment, in which both the patients' health status are diagnosed and proper treatments are provided [7]. Since virtual appointments are provided remotely, they can enhance the delivery of health care to geographically-disadvantaged and medically-underserved populations [1]. In addition, patients who are unable or unwilling to leave their homes to seek medical treatments or are in poor physical condition can also benefit from virtual appointments [4, 8]. Virtual appointments have the potential to enhance primary care delivery by enabling cost reductions for both health delivery and travel and larger panel sizes without sacrifices in the quality of health care [5, 44]. Parallel to its benefits, more patients are willing to receive care through this convenient method. Thus, the demand for virtual appointments is increasing quickly. The total number of virtual consultations is growing by around 10% a year, with growth projected to reach around 25 million in 2020 [49].

Despite the increased usage of virtual appointments and their observed benefits in chronic care, the integration of virtual appointments with office appointments can be operationally challenging for the clinics. One of the reasons for this challenge is that office and virtual appointments can have differences in their treatment/diagnosis effectiveness and in their costs. More specifically, although virtual appointments can provide cost-effective treatments, they can result in similar [16] or worse patient-related outcomes [33, 38] compared to office appointments, which makes it harder to decide how to allocate the available capacity among appointments with different effectiveness. Moreover, with the integration of virtual appointments, the patient flow dynamics become complex and it gets difficult to identify the expected number of patients that can be scheduled for office and virtual appointments. Indeed, faced with rising costs and patient populations, managers of health facilities like clinics strive to determine an appropriate capacity to meet the needs of the patients and avoid the opportunity cost and over-utilization cost as much as possible. Thus, it is important to develop strategies to determine the expected number of patients and allocate available capacity efficiently by considering the patient flow dynamics. To address the need for capacity allocation policies, we study in this paper a chronic care setting in which patients are scheduled for virtual or office appointments. We consider that, similar to office appointments, virtual appointments can also provide both treatment and diagnosis, and parallel to previous studies [7, 33, 38] we assume that virtual appointments can be less effective than office appointments. We develop a modeling framework to determine the optimal allocation of the capacity for both office and virtual appointments and aim to answer the following operational questions:

1. What is the expected number of patients scheduled for office and virtual appointments for the given follow-up, service, arrival, and departure rates?
2. How should the available capacity be allocated among office and virtual appointments to maximize the long-run average earnings of a health clinic?

To address these questions, we develop a migration network model to analyze patient flow and disease progressions. Using the migration network model, we first analytically investigate the number of patients in the steady state who are scheduled for office and virtual appointments. Second, we analyze how the expected number of patients at each node of the migration network would change with respect to some model parameters. Third, we develop a newsvendor-type model to maximize the long-run average earnings of a health clinic. We further propose heuristics to find the capacity allocations among office and virtual appointments. Fourth, we analytically investigate how limited capacity impacts the proposed heuristics and the optimal capacity allocation decisions. Finally, through our numerical studies, we analyze the effect of model parameters on the allocation of the capacity of the office and virtual appointments by analyzing different scenarios.

The remainder of this paper is organized as follows: In Section 2, we review the related literature. Section 3 presents the migration network model and characterizes the number of patients in steady-state conditions. In Section 4, we develop capacity allocation models for different settings and propose heuristics to identify capacity allocations among office and virtual appointments. In Section 5, we perform numerical experiments and sensitivity analysis to illustrate the application of our models. Finally, our conclusions are outlined in Section 6.

2 Relevant literature and contributions

In this section, we discuss separately the relevant literature and the contribution of our study.

2.1 Relevant literature

Our study builds on the literature of decision models in community-based chronic care delivery. Related to this area, [31] present and analyze three representative examples of prevailing quantitative decision models for managing community-based chronic care (i.e., [18, 30]). For each example, they analyze the background of the problem, present the methodology, and show their findings and implications. Among these examples, [30] propose a Markov decision process to model multiple care-provider visit patterns for stroke patients, while [18] combine a Markovian disease progression model with a capacity allocation model to determine revisit intervals for childhood asthma care. A major difference of our paper from the listed literature is that we consider different types of appointments (i.e., office and virtual appointments) and investigate the capacity allocation decisions among the different types of appointments.

Related to the virtual appointment setting, studies that investigate the management of office and virtual appointments are limited. In a relevant study, [36] build an optimization model to design effective checkup plans (i.e., phone calls, office visits) for monitoring individual patients after hospital discharge. Their study considers only the diagnosis impacts of the virtual appointments, whereas we include both the treatment and the diagnosis impact of the virtual appointments. Among the studies considering the impact of virtual appointments on both treatment and diagnosis, [6] develop a Markovian model to determine the patient revisit intervals in primary care by incorporating virtual appointments into an office appointment setting. In another study, [7] develop a stochastic dynamic programming model to determine the follow-up rates for office and virtual appointments, and they investigate the value of virtual appointments in patients' health outcomes. In these papers, the capacity of the appointments is assumed to be given. In contrast to these studies, we investigate the capacity allocation of office and virtual appointments for different settings.

Another stream of literature that is relevant to our study is on the capacity planning problem in health care, which addresses the issue of allocating limited resources to satisfy the demand of the patients. There are several studies in this area, and [26] provide a comprehensive review of resource allocation and capacity planning in health care. Among this literature, the following papers are more relevant to our methodology. [11] develop an optimization/queuing network model for optimal planning of resource allocations (e.g., beds and nurses) and apply it to a blood bank and a health maintenance organization. [32] develop a multi-class migration network model as an optimization model to determine the optimal capacity that maximizes the overall profit of a dialysis clinic. [34] present a long-term care network model to determine the optimal capacity for nursing homes and community-based services. Distinct from the above literature, we consider both patient flows and patients' disease progression to determine optimal capacity allocations. Our differences in the modeling structure are detailed in the following section.

2.2 Contributions

In this paper, we consider an operational problem that deals with the effective allocation of capacity among different types of appointments (i.e., office and virtual appointments) to maximize the long-run average earnings of the clinics. The paper contributes to the capacity allocation literature on the basis of the model structure, since we consider two types of appointments with different effectiveness and patient interactions in the chronic care setting. More specifically, different from the above literature, we consider (i) the integration of operational decisions (i.e., how to allocate capacity among different appointments) with the clinical operations (i.e., treatment and diagnosis of patients), (ii) two types of appointments having different diagnosis and treatment effectiveness, (iii) disease progression due to the chronic nature of the condition, and (iv) patient dynamics and different patient groups rather than homogenous patients, as we categorize patients as controlled vs. uncontrolled and returning vs. new. The differences in the modeling approaches lead us to reach different and unique conclusions. For example, we derive the expected number of patients scheduled for office and virtual appointments and the expected number of patients in the controlled and uncontrolled health states by considering patients' disease progressions. Through our results, we present how office and virtual follow-up rates impact the optimal office and virtual appointment capacity allocations and present how limited total capacity and time impact the allocated capacity and average clinic earnings. We further propose efficient and practical solutions that would help clinics in their capacity allocation decisions and bring them higher average earnings. We find that although virtual appointments are not as effective as office appointments, they can have equal importance to office appointments due to their lower costs. Different from the cited literature, we also show how the expected number of patients at different nodes in the migration network model changes with respect to some of the model parameters.

3 Migration network model for office and virtual appointments

In this section, we consider a cohort of patients receiving chronic care via both office and virtual appointments. In

this network, two types of patients are served (i.e., new patients and returning patients), and physicians provide both office and virtual appointments. We use a continuous-time open migration network [27, p.48-p.57] to simulate the population dynamics (i.e., patient flows and disease progression) in which patients' arrivals are considered as Poisson process and the time intervals between patient transitions are independently and exponentially distributed. We further consider an infinite population so that the node capacities of the migration network are unlimited.

We illustrate our migration network model in Fig. 1, and we describe the nodes and flows of the network in this section. We use $i \in \mathcal{I} = \{o, v\}$, where "$o$" corresponds to office appointments and "v" corresponds to virtual appointments, to denote the set of appointment types. New patients with office and virtual appointments arrive with Poisson arrival rate λ_i, $\forall i \in \mathcal{I}$. We define the "service" as the diagnosis and treatment of a patient, and we consider that office and virtual appointments provide both diagnosis and treatment during the appointment. More specifically, service time corresponds to the duration of an appointment, and service times of patients are exponentially and independently distributed. We use μ_i, $\forall i \in \mathcal{I}$ to denote the service rate of office and virtual appointments, respectively. We define follow-up time (i.e., revisit interval) as the time between the current visit and the next time the patient initiates an appointment. We consider that after each appointment, the physician recommends to the patient the type and the time of the next visit. Hence, patients are scheduled for appointments based on the physician's recommendation. Patient

follow-up times are assumed to be independently and exponentially distributed, with a rate of σ_i, $\forall i \in \mathcal{I}$. Some patients may depart from the physicians' panel before scheduling another appointment (i.e., change the physician or clinic). Patients' departure times are independently and exponentially distributed with a rate of δ.

We use the "control" measure to characterize the patient health status. The "control" measure helps with understanding how well chronic-care related symptoms are currently controlled in a patient. Depending on the types of chronic diseases, these categorizations may differ. For example, for asthma, four categories can be used as follows: (i) controlled, (ii) improved, (iii) unchanged, and (iv) worsened, and the last three are classified as an uncontrolled health state [18]. For the sake of simplification, in our model, we consider two health states as controlled and uncontrolled to characterize the patients' health states. Then, let $j \in \mathcal{J} = \{c, u\}$ represent the set of health states, where "c" corresponds to the controlled health state and "u" corresponds to the uncontrolled health state.

We assume that patients in the network may not be scheduled for an appointment (i.e., may not receive any care) and they may be in the controlled and in the uncontrolled health states. Hence, we define w_c as patients who are in the controlled health state and not scheduled for an appointment, while we use w_u to denote patients who are in the uncontrolled health state and not scheduled for an appointment. We assume that there is no transition from the uncontrolled state to the controlled state without treatment. However, due to disease progression, some of the patients in the controlled
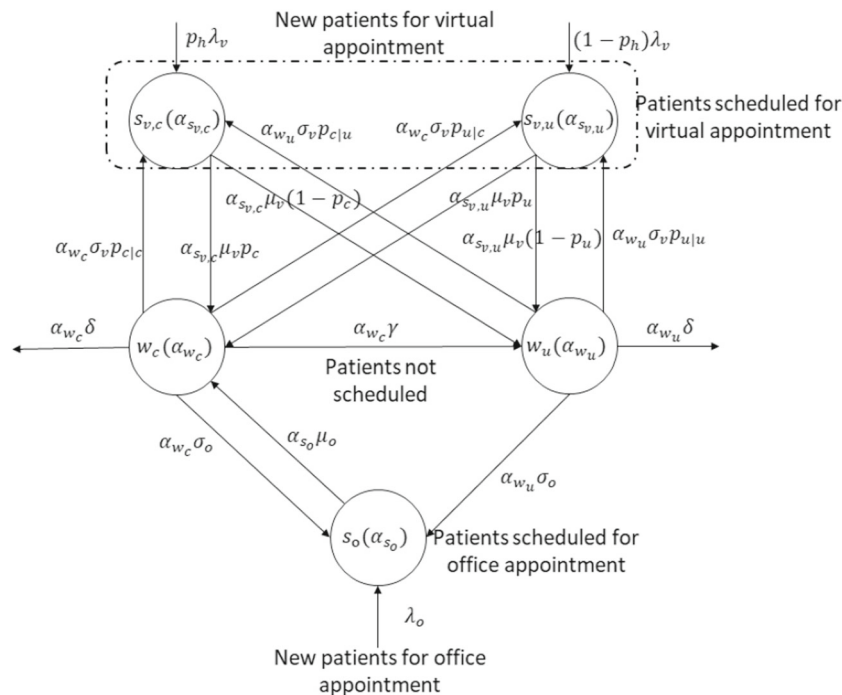

Fig. 1 Migration network model

health state and not receiving care (i.e., w_c) may transition into the uncontrolled health state (i.e., w_u) within the unit time. The time for a controlled patient to progress into the uncontrolled state is assumed to follow an exponential distribution with a rate of γ.

We also consider that patients in the network may be scheduled for office and virtual appointments and receive care. At each type of appointment, the health state of the patient is diagnosed and the patient is treated. We assume that office appointments can be more effective than virtual appointments [7, 33, 38], and the treatment and the diagnosis in the office appointments are perfect, while those of the virtual appointments are imperfect. Perfect treatment means that a patient's health state recovers to the best health state after treatment, while perfect diagnosis means that a patient's health state is revealed accurately during the diagnosis. On the other hand, imperfect treatment means that a patient's health state can transit into a different health state with some probability, while imperfect diagnosis means that a patient's health state may be revealed inaccurately during the diagnosis. The perfect diagnosis/treatment assumption is similar to the ones in the machine maintenance and repair literature as well [9, 43]. Moreover, in the healthcare literature, the perfect diagnosis and treatment assumption is also used by [6, 7, 18]. More specifically, patients in each health state are assumed to be always diagnosed accurately if they are scheduled for an office appointment, and they will be in the controlled health state after the office appointment regardless of their initial health state before the appointment. On the other hand, patients scheduled for a virtual appointment may be diagnosed inaccurately, since virtual appointments are expected to be less precise than office appointments [7, 33, 38].

Thus, we first define s_o to represent patients who are scheduled for office appointments and receiving care (i.e., diagnosis and treatment). Next, to capture the imperfect diagnosis of virtual appointments, we define $s_{v,j}$ to denote patients who are scheduled for virtual appointments and diagnosed in health state $j \in \mathcal{J}$ at the virtual appointment. We use conditional probability to define the imperfect

diagnosis probability for the virtual appointments. Let $p_{j|j'}, \forall j, j' \in \mathcal{J}$ denote the probability that the patient in health state j' is diagnosed in health state j at the virtual appointment, where $p_{c|j'} + p_{u|j'} = 1, \forall j' \in \mathcal{J}$. We also capture the imperfect treatment in virtual appointments. More specifically, a patient diagnosed in the uncontrolled health state at the virtual appointment may remain in the uncontrolled health state with probability $(1 - p_u)$ or may transition into the controlled health state with probability p_u. Similarly, a patient diagnosed in the controlled health state at the virtual appointment may remain in the controlled health state with probability p_c after the virtual appointment or may be in the uncontrolled health state with probability $(1 - p_c)$ after the virtual appointment (since not all patients in the controlled health state may be diagnosed accurately). Hence, the outcomes of nodes $s_{v,c}$ and $s_{v,u}$ are the patients who are being diagnosed and treated during the appointment, and the patients' health statuses may remain the same, may improve, or may get worse after the virtual appointment (i.e., after being diagnosed and treated). We further assume that the new patients scheduled for virtual appointments will be diagnosed in the controlled health state with probability p_h.

Overall, we consider five nodes in the network, and we use $k \in \mathcal{K} = \{w_c; w_u; s_o; s_{v,c}; s_{v,u}\}$ to represent the set of nodes in the migration network. In Fig. 1, we illustrate the described flow of patients between each node through arcs. The arcs between nodes represent the process of a patient that flows from one node to another. For example, the arc from node "s_o" to "w_c" represents the flow of patients from an office appointment to their homes after they have their appointment. We also show the inflow and outflow for each node next to each arc. For example, there are two outflows from node "$s_{v,u}$" where uncontrolled patients can improve to the controlled health state or can remain in the uncontrolled health state after receiving a virtual appointment (i.e., being diagnosed and treated).

We define α_k to denote the expected number of patients at node $k \in \mathcal{K}$ in the steady-state condition. The number of patients at node k satisfy the following balance equations, which are derived from Fig. 1 [27, p.49]:

$$\mu_v \alpha_{s_{v,c}} - \sigma_v p_{c|c} \alpha_{w_c} - \sigma_v p_{c|u} \alpha_{w_u} = p_h \lambda_v \tag{1}$$

$$\mu_v \alpha_{s_{v,u}} - \sigma_v (1 - p_{c|c}) \alpha_{w_c} - \sigma_v (1 - p_{c|u}) \alpha_{w_u} = (1 - p_h) \lambda_v \tag{2}$$

$$-\mu_v p_c \alpha_{s_{v,c}} - \mu_v p_u \alpha_{s_{v,u}} + (\sigma_v + \sigma_o + \delta + \gamma) \alpha_{w_c} - \mu_o \alpha_{s_o} = 0 \tag{3}$$

$$-\mu_v (1 - p_c) \alpha_{s_{v,c}} - \mu_v (1 - p_u) \alpha_{s_{v,u}} - \gamma \alpha_{w_c} + (\sigma_v + \sigma_o + \delta) \alpha_{w_u} = 0 \tag{4}$$

$$-\sigma_o \alpha_{w_c} - \sigma_o \alpha_{w_u} + \mu_o \alpha_{s_o} = \lambda_o \tag{5}$$

These equations represent that the total inflow to node $k \in \mathcal{K}$ must be equal to the total outflow from that node.

Equations 1–5 are the balance equations with five unknowns (i.e., $\alpha_{s_{v,c}}; \alpha_{s_o}; \alpha_{s_{v,u}}; \alpha_{w_c}; \alpha_{w_u}$), and we solve these balance

equations to obtain the average number of patients in each node at steady state. The result for each α_k, $\forall k \in \mathcal{K}$ is included in the Appendix B. By using these equations, we characterize some of the structural properties to show the relationship between the model parameters and the number of patients at each node through Theorem 1.[1]

Theorem 1 *In the migration network model, the number of patients at each specific node presents the following structural properties:*

(a) *Expected number of patients at virtual appointments (i.e., $\alpha_{s_{v,c}} + \alpha_{s_{v,u}}$) is a linearly increasing function of virtual follow-up rate (σ_v), while it is not dependent on the office follow-up rate (σ_o), and expected number of patients at office appointments (i.e., α_o) is a linearly increasing function of office follow-up rate (σ_o), while it is not dependent on the virtual follow-up rate (σ_v).*

(b) *Expected number of patients who are not scheduled for an appointment and in the controlled health state (i.e., α_{w_c}) is an increasing concave function of office follow-up rate (σ_o), and expected number of patients who are not scheduled for an appointment and in the uncontrolled health state (i.e., α_{w_u}) is a decreasing convex function of office follow-up rate (σ_o) if $p_u > p_c$.*

(c) *Expected number of patients who are not scheduled for an appointment and in the controlled health state (i.e., α_{w_c}) is an increasing convex function of p_c, while it is an increasing concave function of p_u. On the other hand, the expected number of patients who are not scheduled for an appointment and in the uncontrolled health state (i.e., α_{w_u}) is a decreasing concave function of p_c, while it is a decreasing convex function of p_u.*

(d) *Expected number of patients who are not scheduled for an appointment and in the controlled health state (i.e., α_{w_c}) is a decreasing convex function of $p_{c|c}$, while it is a decreasing concave function of $p_{c|u}$ if $p_u > p_c$.*

(e) *Expected numbers of patients who are not scheduled for an appointment and in both controlled and uncontrolled health states (i.e., α_{w_c}, and α_{w_u}) are not dependent on the service rate of office and virtual appointments (μ_o and μ_v).*

Theorem 1 (a) shows that as the virtual follow-up rate σ_v (resp. office follow-up rate σ_o) increases (i.e., as patients have virtual visits (resp. office visits) more frequently), the expected number of patients scheduled for virtual appointments (resp. office appointments) increases linearly. On the other hand, office follow-up rate σ_o (resp. virtual follow-up rate σ_v) does not impact the expected number

of patients scheduled for virtual appointments (resp. office appointments). We further investigate how the office follow-up rate impacts the expected number of non-scheduled patients in the controlled and uncontrolled health states. Theorem 1 (b) states that if the condition of $p_u > p_c$ is satisfied, as the office follow-up rate σ_o increases, patients' health statuses are more controlled and the expected number of non-scheduled patients in the controlled health state increases with a concave structure, while the expected number of patients in the uncontrolled health state decreases with a convex structure. Moreover, we investigate the treatment impact of virtual appointments on the expected number of non-scheduled patients in different health states. More specifically, as described in Theorem 1 (c), if the treatment effectiveness of virtual appointments in the controlled and uncontrolled health states (i.e., p_c and p_u) increase, the expected number of non-scheduled patients in the controlled health state (i.e., α_{w_c}) will increase with a convex and a concave structure, respectively. Note that both p_c and p_u can be at most 1, and hence the expected number of non-scheduled patients in the controlled health state will take its highest value when $p_c = p_u = 1$. On the other hand, the increase in the treatment effectiveness of virtual appointments in the controlled and uncontrolled health states (i.e., p_c and p_u) results in a decrease in the expected number of non-scheduled patients in the uncontrolled health state (i.e., α_{w_u}) with a concave and a convex structure, respectively. We also investigate the impact of the diagnosis effectiveness of virtual appointments on non-scheduled patients. In Theorem 1 (d), it is described that as $p_{c|c}$ and $p_{c|u}$ increase, the expected number of non-scheduled patients in the controlled health state (i.e., α_{w_c}) decreases with a convex and a concave structure, respectively. Finally, through Theorem 1 (e) we find that office and virtual service rates (i.e., μ_v and μ_o) do not impact the expected number of non-scheduled patients in both controlled and uncontrolled health states. Overall, through Theorem 1, we show how patient dynamics and flows change at each node with respect to the model parameters.

Next, we use our results from the migration network model to define the capacity allocation model. Hence, we use α_k, $\forall k \in \mathcal{K}$ to define the steady-state distribution π_k, $\forall k \in \mathcal{K}$. Hence, let x_k denote the number of patients at node $k \in \mathcal{K}$. Kelly ([27], p.53) shows that in steady state, the nodes states are independent, and the steady-state distribution for each node is a Poisson distribution with parameter α_k and given by Eq. 6 as follows:

$$\pi_k(x_k = x) = \frac{\alpha_k^x}{x!} \bigg/ \sum_{n=0}^{\infty} \frac{\alpha_k^n}{n!} = e^{-\alpha_k} \frac{\alpha_k^x}{x!}, \quad \forall k \in \mathcal{K} \quad (6)$$

The steady-state distribution defines the probability of having x_k number of patients at each node k. We use these

---

probabilities to define the probabilistic capacity allocation model in Section 4.

4 Capacity allocation model

In this section, we build newsvendor-type capacity allocation models to find the optimal capacity for office and virtual appointments to maximize the long-run average earnings of a clinic. As described in the previous section, we consider that the node capacities of the migration network are unlimited, where the number of patients at each node is unlimited. However, we consider a threshold capacity for office and virtual appointments [34]. The threshold capacity that we assign describes the number of patients that can be served under the regular cost, and the actual number of patients in office and virtual appointments can exceed this threshold capacity. When the number of patients in office and virtual appointments exceeds this threshold capacity, we consider that a penalty cost due to patient overflow occurs. Hence, we aim to find the optimal threshold capacity for office and virtual appointments for the clinics under the assumption that node capacities are unlimited. In the following sections, we first introduce the capacity allocation model without constraints. Then, we modify the unconstrained model by adding constraints on the optimal office and virtual appointment threshold capacities.

4.1 Unconstrained capacity allocation model

We consider a clinic that provides both office and virtual appointments with a threshold capacity of M_{s_o} for office appointments, a threshold capacity of M_{s_v} for virtual appointments, and the total threshold capacity of $M = M_{s_o} + M_{s_v}$. Since in our migration network we split the virtual appointments into two parts to reflect the imperfect diagnosis and treatment, we define $M_{s_{v,c}}$ and $M_{s_{v,u}}$, which denote the threshold capacity for virtual appointments in the controlled and uncontrolled health states, respectively (i.e., $M_{s_v} = M_{s_{v,c}} + M_{s_{v,u}}$). Defining different types of capacities for virtual appointments ensures more flexibility in the model definition and our findings can impact the following two areas in practice: (i) capacity splitting decisions and (ii) patient scheduling decisions. In practice, although it is not common to split the capacity of virtual appointments according to the patients' health statuses, there are different implementations of virtual appointments and especially after COVID-19, they have been used in practice more often. More specifically, (i) some virtual appointments can be provided asynchronously for patients in controlled

health states [39], (ii) some of them can be provided synchronously for both controlled and uncontrolled patients [28], and (iii) some of them can be provided synchronously for emergency patients (i.e., tele-emergency) [20]. Hence, clinics can consider splitting their virtual appointment capacities based on patient needs. On the other hand, clinics do not need to split their virtual appointment capacity, but they can implement the proposed policies in their patient scheduling decisions. More specifically, during the appointments, physicians diagnose patients, provide treatments, and by considering the patient's overall health status and progress, the provider recommends the next appointment time to the patient. Thus, when physicians are making their scheduling recommendations for the next appointment they can take into account the capacity allocation decisions (i.e., the ratio of $M_{s_{v,c}}/M_{s_{v,u}}$).

Let $\mathcal{K}' = \{s_o; s_{v,c}; s_{v,u}\}$ be the set of updated migration network nodes. Then, we first define $r_k \ \forall k \in \mathcal{K}'$ as the marginal profit of office and virtual appointments. We note that there is one type of virtual appointment and $r_{s_v} = r_{s_{v,c}} = r_{s_{v,u}}$ (i.e., the difference between the revenue and variable cost[2] is the same for all virtual appointments). Similarly, we define $\eta_k, \forall k \in \mathcal{K}'$ to denote the unit threshold capacity cost for office and virtual appointments, where unit threshold capacity cost is the fixed cost of allocating capacity, which can be employee salaries, building-related costs, and equipment. We assume that $r_k > \eta_k$ [32]. By assuming $r_k > \eta_k$, we ensure that the optimal threshold capacity M_k is greater than 0. More specifically, if the unit threshold capacity cost is larger than or equal to the marginal profit, it will be optimal to provide no service and $M_k = 0$. Since there is one type of virtual appointment, we consider that $\eta_{s_v} = \eta_{s_{v,c}} = \eta_{s_{v,u}}$. We further assume that the number of patients at the office and virtual appointments can exceed the allocated threshold capacity, and in this case, the clinic provides the corresponding appointment but at a higher total cost. To reflect the cost of patient overflow, we define $f_k > 0, \forall k \in \mathcal{K}'$, which represents the unit net penalty cost of the overflow, where $f_{s_v} = f_{s_{v,c}} = f_{s_{v,u}}$. The definition is similar to the definition of the overbooking cost used by [32]. It is the net cost of meeting the overflow demand, which is the difference between the total variable cost of meeting the extra demand and the revenue earned for that appointment. The clinic still earns the marginal profit r_k for the overflow patients, but the extra variable cost of meeting this excess demand is more than the marginal profit. Let $x_k(t)$ denote the current number of patients at

---

[2]Variable costs are the type of costs that can change depending on the number of patients served, such as hourly labor cost or the cost of materials and supplies.

node k, $\forall k \in \mathcal{K}'$ at time t. Then, our unconstrained capacity allocation model can be defined as follows:

$$max \quad A(M) = \lim_{T \to \infty} \frac{1}{T} \left\{ \sum_{k \in \mathcal{K}'} \int_0^T r_k min\,[x_k(t), M_k]\,dt \right.$$
$$- \sum_{k \in \mathcal{K}'} \int_0^T \eta_k M_k dt$$
$$\left. - \sum_{k \in \mathcal{K}'} \int_0^T f_k\,(x_k(t) - M_k)^+\,dt \right\} \quad (7)$$
$$s.t.$$
$$M_k \geq 0 \qquad\qquad \forall k \in \mathcal{K}' \quad (8)$$

As noted before, the objective function is defined as the function of threshold capacity and the number of patients at each node. Since the number of patients at each node is uncertain, we use the steady-state probabilities π_k defined in Section 3. In objective function (7), the first term represents the marginal profit generated from office and virtual appointments, the second term represents the fixed threshold capacity cost, and the third term represents the penalty costs associated with a capacity shortage. Equation 8 defines the non-negativity setting.

Let $\mathbb{E}_{\pi_k}(x_k)$, $k \in \mathcal{K}'$ be the expected number of patients at node k under the steady-state distribution π_k, $k \in \mathcal{K}'$. Due to the ergodicity of the open migration network [27, p.49], we can define the following equations:

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T x_k(t)dt = \mathbb{E}_{\pi_k}(x_k) = \alpha_k \quad (9)$$

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T [M_k - x_k(t)]^+\,dt = \mathbb{E}_{\pi_k}\big[(M_k - x_k)^+\big] \quad (10)$$

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T [x_k(t) - M_k]^+\,dt = \mathbb{E}_{\pi_k}\big[(x_k - M_k)^+\big] \quad (11)$$

Then, we reformulate the objective function (7) with the following equation, and we include the detailed steps of the reformulation in the Appendix B.

$$A(M) = \sum_{k \in \mathcal{K}'} \left[(r_k - \eta_k)\alpha_k - \eta_k \mathbb{E}_{\pi_k}\big[(M_k - x_k)^+\big] \right.$$
$$\left. - (f_k + r_k - \eta_k)\mathbb{E}_{\pi_k}\big[(x_k - M_k)^+\big] \right] \quad (12)$$

Similar to Eq. 7, in Eq. 12, the first term is the difference between the marginal profit and the fixed cost. The second term is the opportunity cost for unutilized capacity, and the last term represents the cost due to patient overflow.

We define $A_k(M_k)$ as the individual objective function for appointment $k \in \mathcal{K}'$, and it can be defined as follows:

$$A_k(M_k) = (r_k - \eta_k)\alpha_k - \eta_k \mathbb{E}_{\pi_k}\big[(M_k - x_k)^+\big] \quad (13)$$
$$- (f_k + r_k - \eta_k)\mathbb{E}_{\pi_k}\big[(x_k - M_k)^+\big], \qquad \forall k \in \mathcal{K}'$$

To maximize the objective $A(M)$, each sub-objective $A_k(M_k)$ can be maximized separately. We derive the optimal threshold capacity M_k, $\forall k \in \mathcal{K}'$ for this unconstrained capacity planning model through Proposition 1.

Proposition 1 *The optimal solution of the unconstrained capacity allocation model, denoted by $M^* = (M_{s_o}^{min}, M_{s_{v,c}}^{min}, M_{s_{v,u}}^{min})$, is given by*

$$M_k^{min} = min\left\{ M_k \geq 0 : \pi_k(x_k \leq M_k) \geq \frac{f_k + r_k - \eta_k}{f_k + r_k} \right\}$$
$$\forall k \in \mathcal{K}' \quad (14)$$

where $\pi_k(x_k \leq M_k)$ is the cumulative probability that x_i is less than or equal to M_k, $\forall k \in \mathcal{K}'$.

According to Eq. 14, the optimal threshold capacity depends on both Poisson distribution parameter α_k and the ratio of $\frac{f_k + r_k - \eta_k}{f_k + r_k}$. For a fixed Poisson parameter α_k, the cumulative probability function is a non-decreasing function of M_k. Hence, as the ratio $\frac{f_k + r_k - \eta_k}{f_k + r_k}$ increases, the optimal threshold capacity will also increase.

4.2 Capacity allocation model with capacity constraint

In practice, due to the limited resources of the clinic, the number of regular office and virtual appointments may be limited. In this section, we extend the unconstrained capacity allocation model presented in Section 4.1 and investigate the impact of adding an upper bound on the optimal threshold capacity decisions. This change does not impact the balance equations of the migration network and patient flows, and the objective function $A(M)$ remains the same as with the unconstrained model. More specifically, node capacities in the migration network are unlimited, and it is still allowed to have more than the M_k number of patients [34]. Hence, Eqs. 1–6 still hold when the constraint (16) is added. We use TC to denote the limited total threshold capacity. Then, the capacity allocation model can be updated as follows:

$$max \quad A(M) = \sum_{k \in \mathcal{K}'} \left[(r_k - \eta_k)\alpha_k - \eta_k \mathbb{E}_{\pi_k}\big[(M_k - x_k)^+\big] \right.$$
$$\left. - (f_k + r_k - \eta_k)\mathbb{E}_{\pi_k}\big[(x_k - M_k)^+\big] \right] \quad (15)$$
$$s.t. \sum_{k \in \mathcal{K}'} M_k \leq TC \quad (16)$$
$$M_k \geq 0 \qquad\qquad \forall k \in \mathcal{K}' \quad (17)$$

In the model, Eq. 16 states that the allocated threshold capacity should be less than or equal to the total available threshold capacity TC, and Eq. 17 defines non-negativity constraints. Let M_{TC} be the allocated threshold capacity when the total threshold capacity is limited. Recall that

$M^* = (M^{min}_{s_o}, M^{min}_{s_{v,c}}, M^{min}_{s_{v,u}})$ is the optimal threshold capacity for the unconstrained capacity allocation model given in Proposition 1. It is clear that if $\sum_{k \in \mathcal{K}'} M^{min}_k \leq TC$, then $M_{TC} = M^*$. This means that the clinic has enough resources, which maximizes its overall average earnings, and the clinic may be considered not to have excess capacity. If $\sum_{k \in \mathcal{K}'} M^{min}_k \geq TC$, this means that the clinic has scarce resources, and the optimal solution provided in Proposition 1 may not hold. Hence, we propose Heuristic 1 to determine the capacity allocation decisions. Let M^t_k, $k \in \mathcal{K}'$ be the threshold capacity of node k at iteration t. Heuristic 1 uses the partial derivation calculation of the objective function, so we define the partial derivative of the objective function for each $k \in \mathcal{K}'$ as follows:

$$A'(M_k) = \frac{\Delta A(M)}{\Delta M_k} = A(M_k + 1) - A(M_k) = f_k + r_k$$
$$- \eta_k - (f_k + r_k)\pi_k(x_k \leq M_k) \quad \forall k \in \mathcal{K}' \quad (18)$$

Then, Heuristic 1 can be stated as follows:

Algorithm 1 Capacity allocation heuristic for the model with equal-weighted capacity constraint.

1: $t = 0$ and $M^t_k \leftarrow 0$ for any $k \in \mathcal{K}'$
2: **while** $\sum_{k \in \mathcal{K}'} M^t_k < TC$ **do**
3:     Calculate $A'(M^t_k), \forall k \in \mathcal{K}'$
4:     **if** $A'(M^t_k) \leq 0, \forall k \in \mathcal{K}'$ **then**
5:         **break**
6:     **end if**
7:     $x = argmax_k \, A'(M^t_k), \forall k \in \mathcal{K}'$
8:     $M^{t+1}_x \leftarrow M^t_x + 1$ and $t \leftarrow t + 1$
9:     $M^t_{TC} \leftarrow (M_{s_o}{}^t, M^t_{s_{v,c}}, M^t_{s_{v,u}})$
10: **end while**
11: $M_{TC} \leftarrow M^t_{TC}$

In Heuristic 1, we calculate the marginal gain of having one more unit of the threshold capacity for the office and virtual appointments. At each step, we compare the marginal gain of having one office and one virtual appointment and increase the threshold capacity of the appointment with the highest gain by one. The heuristic stops when the allocated capacity reaches the available total threshold capacity or when adding one more unit of threshold capacity for all appointments yields a negative profit gain.

4.3 Capacity allocation model with time constraint

In this section, we take into account the total time required for providing each type of appointment. Section 4.2 assumes that office and virtual appointments both take an equal amount of time. However, virtual appointments are expected

to be shorter than office appointments. Hence, we update (16) by considering the total available time and service times of office and virtual appointments. Let T_w denote the average total available time for the clinic. As defined in Section 4.1, $\mu_i, \forall i \in \mathcal{I}$ represents the service rate of an office and a virtual appointment, respectively. We assume that virtual appointments for controlled and uncontrolled patients have the same service rate, where $\mu_{s_{v,c}} = \mu_{s_{v,u}} = \mu_v$. Hence, $\frac{1}{\mu_i}, \forall i \in \mathcal{I}$ represents the average service time of appointment type i. Then the capacity allocation model with the time constraint can be defined as follows:

$$max \quad A(M) = \sum_{k \in \mathcal{K}'} \left[(r_k - \eta_k)\alpha_k - \eta_k \mathbb{E}_{\pi_k}\left[(M_k - x_k)^+\right] \right.$$
$$\left. - (f_k + r_k - \eta_k)\mathbb{E}_{\pi_k}\left[(x_k - M_k)^+\right] \right] \quad (19)$$
$$s.t. \quad \frac{1}{\mu_o}M_{s_o} + \frac{1}{\mu_v}(M_{s_{v,c}} + M_{s_{v,u}}) \leq T_w \quad (20)$$

where the objective function remains the same. Let M_{T_w} be the threshold capacity allocation decision for the model with limited time. Recall that $M^* = (M^{min}_{s_o}, M^{min}_{s_{v,c}}, M^{min}_{s_{v,u}})$ is the optimal threshold capacity for the unconstrained capacity allocation model given in Proposition 1. Similar to Section 4.2, if $\frac{1}{\mu_o}M_{s_o}{}^{min} + \frac{1}{\mu_v}(M^{min}_{s_{v,c}} + M^{min}_{s_{v,u}}) \leq T_w$, then $M_{T_w} = M^*$. On the other hand, if the resources are scarce, Proposition 1 does not hold. Hence, we define Heuristic 2 to determine the capacity allocation decisions when the total time is limited. Let $M^t_k, k \in \mathcal{K}'$ be the threshold capacity of node k at the tth iteration. We further define $Z_k(M_k) = A'(M_k)/\frac{1}{\mu_k} = \mu_k A'(M_k)$, which is the marginal profit gain of node $k \in \mathcal{K}'$ when total available time is limited.

Similar to Heuristic 1, Heuristic 2 also compares the marginal gains of different types of appointments at each iteration and increases the threshold capacity of the appointment with the highest positive gain by 1. Neither Heuristic 1 nor Heuristic 2 guarantees the optimal solution. Hence, we analyze the relative errors of the solutions of the heuristics from the optimal solution in the following proposition:

Proposition 2 *The relative errors of the solutions of Heuristic 1 (i.e., M_{TC}) and Heuristic 2 (i.e., M_{T_w}), which are the approximations of M^*_{TC} and $M^*_{T_w}$, respectively, are no greater than the relative error terms of $\frac{max_k \, A'(M^t_k)}{A(M_{TC})}$ and $\frac{max_k \, A'(M^t_k)}{A(M_{T_w})}$, respectively. We state the corresponding equations as follows:*

$$\frac{A(M^*_{TC}) - A(M_{TC})}{A(M^*_{TC})} \leq \frac{A(M^*_{TC}) - A(M_{TC})}{A(M_{TC})} < \frac{max_k \, A'(M^t_k)}{A(M_{TC})} \quad (21)$$
$$\frac{A(M^*_{T_w}) - A(M_{T_w})}{A(M^*_{T_w})} \leq \frac{A(M^*_{T_w}) - A(M_{T_w})}{A(M_{T_w})} < \frac{max_k \, A'(M^t_k)}{A(M_{T_w})} \quad (22)$$

where M_k^t represents the threshold capacity in the final iteration of the heuristics. Also, $A'(M_k^t)$ is the marginal gain of appointment type $k \in \mathcal{K}'$ in the final iteration. Proposition 2 ensures that the percent profit gap between the optimal solution and the proposed heuristic solutions is not greater than the percent marginal profit gain in the final iteration.

Algorithm 2 Capacity allocation heuristic for the model with limited time constraint.

1: $t = 0$ and $M_k^t \leftarrow 0, \forall k \in \mathcal{K}'$
2: **while** $\frac{1}{\mu_o} M_{s_o}^t + \frac{1}{\mu_v}(M_{s_{v,c}}^t + M_{s_{v,u}}^t) \leq T_w$ **do**
3: $\quad Z_k(M_k^t) = \mu_k A'(M_k^t), \forall k \in \mathcal{K}'$
4: $\quad$ **if** $\sum_{k \in \mathcal{K}'} \frac{1}{\mu_k} M_k^t + \frac{1}{\mu_{k'}} > T_w, \forall k' \in \mathcal{K}'$ **then**
5: $\quad\quad$ **break**
6: $\quad$ **end if**
7: $\quad x = argmax_k\ Z_k(M_k^t), \forall k \in \mathcal{K}'$
8: $\quad M_x^{t+1} \leftarrow M_x^t + 1$ and $t \leftarrow t + 1$
9: $\quad M_{T_w}^t \leftarrow (M_{s_o}^t, M_{s_{v,c}}^t, M_{s_{v,u}}^t)$
10: **end while**
11: $M_{T_w} \leftarrow M_{T_w}^t$

---

5 Numerical studies

In this section, we perform numerical experiments to analyze how the optimal capacity allocation decision varies under different scenarios. To this end, we first describe the model parameter estimation process. Then, we investigate the change in the optimal threshold capacity with respect to the follow-up rate and capacity constraint. Finally, we compare the proposed solutions with some common policies in practice.

5.1 Parameters estimation

In this section, we describe how the model parameters are obtained. We note that our data are based on the literature, and we use several sources to find the parameter values. The parameter values we obtain represent different characteristics. Due to the variation in the parameters' characteristics, fluctuation in the results can be expected. The parameters that we present in this section is our initial setting, and the results that are obtained based on a single setting may not be generalized. To overcome these issues, in Sections 5.2 and 5.3, we define several scenarios and investigate the changes in the proposed office and virtual appointment capacities by considering possible fluctuations in the parameter values. We present minimum, average, and maximum values of the proposed threshold capacity values to provide a range for the decision-makers.

Flow parameters Based on a survey of American physicians [45], doctors see 20.2 patients per day on average, and physicians work on average 51.40 hours per week (including all clinical and non-clinical duties). Of these, physicians work on average 11.37 hours per week on non-clinical (paperwork) duties only. Hence, we calculate the average service time for each patient as $(51.4 - 11.37)/(5 \times 20.2) = 0.396$ hours and the service rate of the office appointment as $\mu_o = 1/0.396 = 2.525$/hour. In addition, the average appointment time of the virtual appointments is less than that of the office appointments, and it is reported as around 12 minutes [47]. Thus, the virtual service rate is estimated as $\mu_v = \mu_{s_{v,c}} = \mu_{s_{v,u}} = 60/12 = 5$/hour. To calculate the new patient arrival rate, we consider the state of Michigan. The population of Michigan was 9.976 million in 2017, and 47.9% of them suffered from chronic diseases [41], while in 2018, the Michigan population increased to 9.996 million, and 48.1% of them suffered from chronic diseases [46]. Thus, the number of new chronic patients in Michigan can be calculated as $(9.996 \times 48.1\% - 9.976 \times 47.9\%) \times 10^6 = 29,572$. There are 278 clinics in the state of Michigan [15], from which number the total monthly new arrival rate may be estimated as $\lambda_o + \lambda_v = 29572/(278 \times 12) = 8.865$ patients/month. Also, around 10.4% percent of the visits occur through virtual appointments [45]. Thus, $\lambda_o = 8.865 \times (1 - 10.4\%) = 7.943$ patients/month, $\lambda_v = 8.865 - 7.943 = 0.922$ patients/month. According to the regulations on chronic care management (CCM) [19], a patient should receive at least 20 minutes of clinical care by a physician or other qualified health care professional per calendar month. Considering the service rates of office and virtual appointments, the follow-up rate can be estimated as $\sigma_o = \frac{20 minutes/month}{0.396 hour} = 0.842$/month, and $\sigma_v = \frac{20 minutes/month}{12 minutes} = 1.667$/month. Based on a CDC report [14], generally incurable and ongoing chronic diseases affected approximately 133 million Americans in 2009, representing more than 40% of the total population in this country. In 2009, 7 out of 10 deaths in the U.S. were due to chronic diseases, and the death rate due to chronic diseases was around 1.706 million [29]. Thus, the monthly departure (i.e., death) rate can be calculated as $\delta = 1.706/(133 \times 12) = 0.00107/month$. In addition, the disease progression (i.e., transferring from a controlled health state to an uncontrolled health state) is estimated as $\gamma = 0.5/week$ [13]. To estimate the new patients' health status, we consider a study that considers diabetes patients. [17] present that 43% of new patients among their analytic sample had indications of uncontrolled diabetes. Hence, we assume $p_h = 1 - 43\% = 57\%$, which is the probability that the newly arrived patient in a virtual appointment is in the controlled health state. We note that for the remaining parameters (p_c, p_u, $p_{c|c}$, and $p_{c|u}$) we perform

sensitivity analysis to investigate their effects on the capacity allocation.

Revenue and cost We assume that the workday for a clinic is 20 days per month. For the revenue and cost parameters of the office appointment, we refer to the study of [32]. Then, the marginal profit of an office appointment is estimated as r_{s_o} = \$131 /day ×20 days/month= \$2620/month, the fixed threshold capacity cost of an office appointment is estimated as η_{s_o} = \$84.6/day×20days/month = \$1692/month, and the penalty cost of an office appointment is estimated as f_{s_o} = \$50/day ×20 days/month = \$1000/month [32]. For the virtual appointments, there is no direct historical data, but the total cost of virtual appointments is estimated as 32% less than that of traditional office appointments [2]. Hence, the fixed threshold capacity cost for virtual appointments is estimated as $\eta_{s_{v,c}}$ = $\eta_{s_{v,u}}$ = \$1692/month ×(1−32%) = \$1150.56/month. For the marginal profit and the penalty cost of the virtual appointments, we use the same parameter values as with the office appointments (i.e, $r_{s_{v,c}}$ = $r_{s_{v,u}}$ = r_{s_o}, and $f_{s_{v,c}}$ = $f_{s_{v,u}}$ = f_{s_o}). Table 1 summarizes the value of the patient-flow and profit-related parameters together with the sources from which they are estimated. We note that the parameter values listed in Table 1 are used in one of the scenarios. Then, we analyze several scenarios by considering the possible fluctuations in the parameter values. Hence, through our scenarios we also consider the cases where $r_{s_{v,c}}$ = $r_{s_{v,u}}$ > r_{s_o} or $r_{s_{v,c}}$ = $r_{s_{v,u}}$ < r_{s_o} (resp. where $f_{s_{v,c}}$ = $f_{s_{v,u}}$ > f_{s_o} or $f_{s_{v,c}}$ = $f_{s_{v,u}}$ < f_{s_o}).

5.2 Impact of the model parameters on the threshold capacity

In this section, we investigate the changes in the optimal threshold capacity and average earnings as some of the key parameters in the model change. The parameters we use in the model are obtained from the literature and are not specific to any healthcare organization. For that reason, we perform sensitivity analysis to investigate the optimal threshold capacity for varying parameter values to ensure that the changes in the parameter values due to the different clinics' characteristics can be addressed. We show how the optimal threshold capacity can change with respect to a change in other parameter values. Through our results, we show not only how the optimal threshold capacity changes but also the range of the change in optimal threshold capacity and average earnings. Our sensitivity analysis would also help us to generalize our results for a varying set of parameters and data sets.

Follow-up rate We first study the impact of follow-up rate (i.e., σ_o, σ_v) on the optimal threshold capacity. It has important relevance since reducing or increasing the follow-up rate implies a lower or greater frequency of patients' visits. We vary the follow-up rate in a range between 0.5σ and 1.5σ and present the corresponding optimal capacities in Fig. 2. The results show that as we increase the follow-up rate of office appointments, the optimal threshold capacity of office appointments increases monotonically, while the

Table 1 List of model parameters

| Parameters	Values	Sources
Office service rate (μ_o)	2.525/hour	[45]
Virtual service rate (μ_v)	5/hour	[47]
Office arrival rate (λ_o)	7.943 patients/month	[41, 46],
Virtual arrival rate (λ_v)	0.922 patients/month	[45]
Office follow-up rate (σ_o)	0.842/month	[19]
Virtual follow-up rate (σ_v)	1.667/month	
Departure/death rate (δ)	0.00107/month	[14, 29]
Disease progression rate(γ)	0.5/week	[13]
Probability that the new arriving patient is in the controlled health state (p_h)	0.57	[17]
Probability that patients diagnosed in the controlled health state remain in the controlled health state (p_c)	0.9	
Probability that patients diagnosed in the uncontrolled health state improve to a controlled health state (p_u)	0.7	
Probability controlled patient is diagnosed as controlled ($p_{c|c}$)	0.9	
Probability uncontrolled patient is diagnosed as controlled ($p_{c|u}$)	0.2	
Office marginal profit (r_{s_o})	\$2620/month	
Virtual marginal profit ($r_{s_{v,c}}, r_{s_{v,u}}$)	\$2620/month	
Office overflow penalty cost (f_{s_o})	\$1000/month	[32]
Virtual overflow penalty cost ($f_{s_{v,c}}, f_{s_{v,u}}$)	\$1000/month	
Office fixed threshold capacity cost (η_{s_o})	\$1692/month	
Virtual fixed threshold capacity cost ($\eta_{s_{v,c}}, \eta_{s_{v,u}}$)	\$1150.56/month	[2]
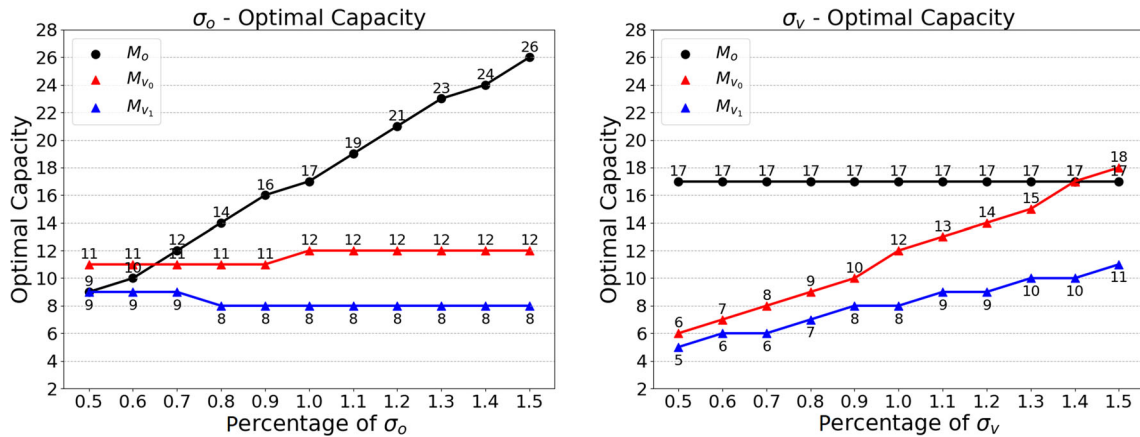
Fig. 2 The impact of follow-up rate on the optimal threshold capacity

total optimal threshold capacity for virtual appointments does not change. This is reasonable, because the increase in the office follow-up rate would result in an increase in the expected number of patients in the office appointments at the steady state but not in the expected number of patients in the virtual appointments. It is also observed that the expected number of patients in virtual appointments is not a function of the office follow-up rate as stated in Eq. 39 in the Appendix B. There occurs a slight increase in the optimal threshold capacity of controlled patients, the reason for which is that the increasing follow-up rate of office appointments transfers more patients from the uncontrolled health state into the controlled health state. Similarly, as the follow-up rate of virtual appointments increases, it is observed that the optimal threshold capacity for controlled health state increases more than the optimal threshold capacity for uncontrolled health state. Although in practice there may be no distinction between capacity for uncontrolled and controlled conditions, this an indication that the increasing follow-up rate of virtual appointments relatively reduces the number of patients in the uncontrolled health state. Thus, increasing the follow-up rate of either office or virtual appointments can improve the health condition of the patients, but it simultaneously increases the demand for office and virtual appointments, which is a challenge for the clinic. This result is consistent with the practice that as the average follow-up rate of the patients increases, the panel size of one physician would decrease. Thus, to serve the same number of patients, more physicians are needed for the clinic.

Limited capacity Next, we analyze the impact of limited total threshold capacity (TC) on the optimal threshold capacity and the average earnings of the clinic. To this end, we first determine M^*, which is the optimal threshold capacity allocation vector for the unconstrained model. For the given model parameters in the unconstrained model,

the optimal threshold capacity for office appointments, for controlled virtual appointments, and for uncontrolled virtual appointments is 17, 12, and 8, respectively, and the optimal total threshold capacity of the clinic is 37. Hence, we change the range of limited total threshold capacity (TC) from 25 to 40. As shown in Fig. 3, as the limited total threshold capacity increases, the office and virtual capacities increase until they reach the optimal. We can see that when the limited total threshold capacity (TC) is greater than 37 (i.e., $\sum_{k \in \mathcal{K}'} M_k^{min} \leq TC$), the threshold capacity allocation vector for the model with the equal-weighted capacity constraint (i.e., M_{TC}) is equal to the optimal result of the unconstrained model (i.e., M^*), which is consistent with the analysis in Section 4.2.

In addition, as we increase the limited total threshold capacity, the average long-run earnings of office and virtual appointments increase and the marginal profit gain decreases, which is consistent with Eq. 87, that the second-order derivation of the long-run average earnings is a monotonic decreasing function.

Limited time We also study the impact of limited working time (T_w) on the threshold capacity and the average earnings of the clinic. Recall that the optimal threshold capacity for the unconstrained model is $M^* = \{M_{s_o}^{min} = 17, M_{s_{v,c}}^{min} = 12, M_{s_{v,u}}^{min} = 8\}$. By considering the service rates (i.e., $\mu_o = 2.525/hour$ and $\mu_v = 5/hour$), the optimal threshold capacity for the unconstrained model is around 11 hours. Hence, we vary the range of the limited working time from 8 to 12 hours. Then, we use Heuristic 2 to obtain the allocation decision for this problem. As shown in Fig. 4, when $T_w = 8\ hours$, the allocated capacities for office and virtual appointments for controlled and uncontrolled patients are 11, 11, and 7, and when $T_w = 10\ hours$, the allocated capacities for office and virtual appointments for controlled and uncontrolled patients are 16, 11, and 7.
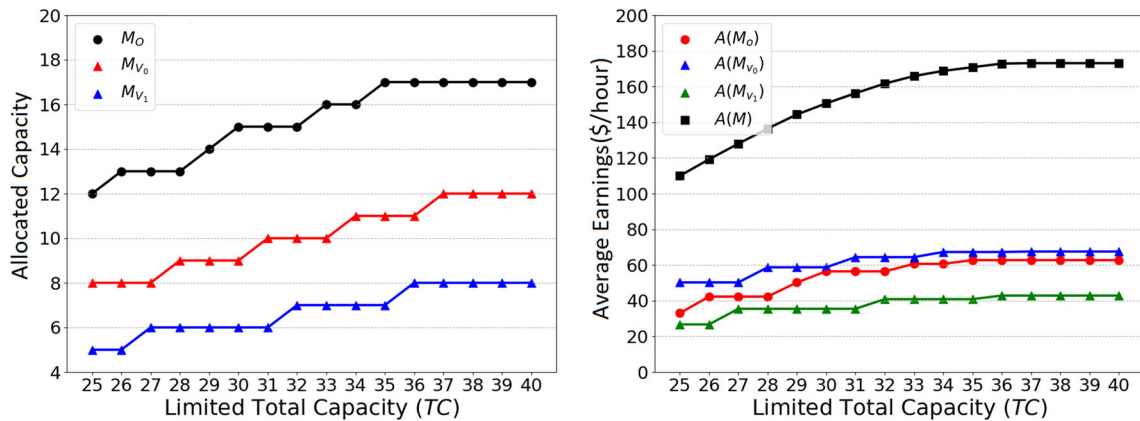
Fig. 3 The impact of limited threshold capacity (TC) on the optimal threshold capacity and average earnings

We observe that the change in the limited time affects the threshold capacity for office appointments more than the threshold capacity for virtual appointments. This is because the average service time of office appointments is nearly twice that of virtual appointments. If the limited time decreases, it becomes more profitable to reduce the threshold capacity for office appointments by one unit rather than reducing the threshold capacity for virtual appointments by two units. As shown, if the limited time is greater than 11 hours, the actual working time remains constant, which is consistent with the analysis in Section 4.2. Since $\frac{1}{\mu_o} M_{s_o}{}^{min} + \frac{1}{\mu_v}(M_{s_{v,c}}^{min} + M_{s_{v,u}}^{min}) \le T_w$, $M_{T_w} = M^*$.

5.3 Comparison of policies

In this section, we compare the total profits of some common benchmark policies with those of our proposed policies (i.e., optimal policy, Heuristics 1 and 2). As benchmark policies, we consider three varying ratios of office appointment threshold capacity to virtual appointment threshold capacity (i.e., M_{s_o}/M_{s_v}): (i) Policy-1: $M_{s_o}/M_{s_v} = 2$, (ii) Policy-2: $M_{s_o}/M_{s_v} = 1$, and (iii) Policy-3: $M_{s_o}/M_{s_v} = 1/2$. Initially, we consider that the virtual appointment capacities allocated for controlled and uncontrolled patients are equal to each other. For comparison, we analyze several scenarios

by varying the parameter values. As the number of varying parameters increases, the number of scenarios and the complexity of the analyses increase. Hence, considering that the impact of the δ and γ variables on the capacity allocation decisions can be small and that preserving the relationship of $\mu_o \le \mu_v$ is important, we keep these variables constant. For all 16 remaining parameters, we use two possible values (i.e., low, high). We use the following formulas to calculate the low and high levels for each parameter:

$$\text{Low Value of a Parameter} = (1 - \text{Fluctuation Rate})$$
$$\times \text{Original Parameter Value} \tag{23}$$
$$\text{High Value of a Parameter} = (1 + \text{Fluctuation Rate})$$
$$\times \text{Original Parameter Value} \tag{24}$$

By considering all possible combinations, we evaluate $2^{15} = 32,768$ scenarios for the unconstrained model, and for the model with the capacity and time constraint, we analyze $2^{16} = 65,536$ scenarios as we also change the parameter TC and T_w. In Tables 2 and 3, we present the solutions obtained from the proposed heuristics (i.e., the ratio of M_{s_o}/M_{s_v}, $M_{s_{v,c}}/M_{s_{v,u}}$, and $A(M)$) and the comparison results of the policies (i.e., Policy-1, Policy-2, and Policy-3) with respect to the proposed heuristics for the unconstrained model, the capacity-constrained model, and
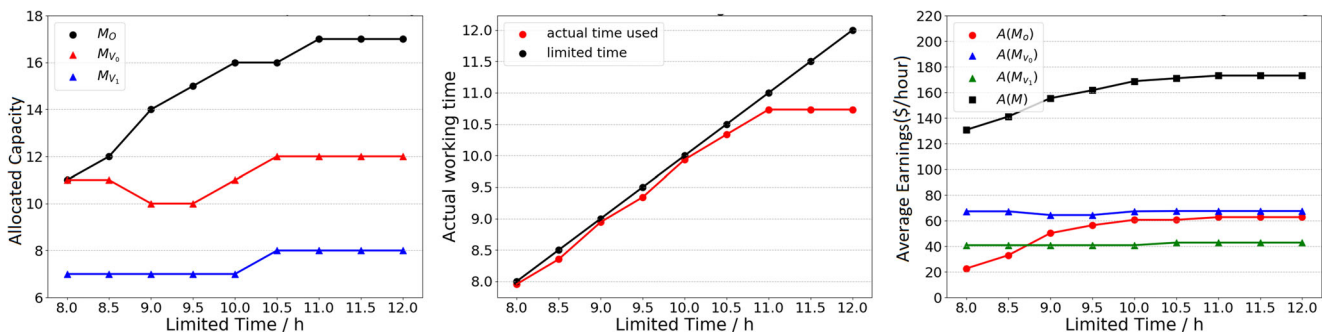

Fig. 4 The impact of limited working time (T_w) on capacity allocation and average earnings

Table 2 Comparison of Policy-1, Policy-2, and Policy-3 with benchmark policies when the fluctuation rate is 5%

	Optimal						% Profit Gap		
Unconstrained Model	M_{s_o}	M_{s_v}	$\frac{M_{s_o}}{M_{s_v}}$	$M_{s_{v,c}}$	$M_{s_{v,u}}$	$\frac{M_{s_{v,c}}}{M_{s_{v,u}}}$	Policy-1	Policy-2	Policy-3
Average	17.45	19.75	0.89	11.59	8.16	1.43	47.61%	5.12%	28.60%
Max	20.00	23.00	1.11	14.00	10.00	1.86	83.31%	18.31%	53.08%
Min	15.00	17.00	0.71	9.00	6.00	1.11	25.40%	0.00%	12.81%
	Heuristic-1						% Profit Gap		
Model with Capacity Constraint	M_{s_o}	M_{s_v}	$\frac{M_{s_o}}{M_{s_v}}$	$M_{s_{v,c}}$	$M_{s_{v,u}}$	$\frac{M_{s_{v,c}}}{M_{s_{v,u}}}$	Policy-1	Policy-2	Policy-3
Average	15.28	16.71	0.92	9.94	6.76	1.49	43.25%	6.20%	27.94%
Max	18.00	20.00	1.14	12.00	9.00	2.20	82.79%	24.33%	53.14%
Min	13.00	14.00	0.70	8.00	5.00	1.11	16.31%	0.00%	7.93%
	Heuristic-2						% Profit Gap		
Model with Time Constraint	M_{s_o}	M_{s_v}	$\frac{M_{s_o}}{M_{s_v}}$	$M_{s_{v,c}}$	$M_{s_{v,u}}$	$\frac{M_{s_{v,c}}}{M_{s_{v,u}}}$	Policy-1	Policy-2	Policy-3
Average	16.50	19.25	0.86	11.30	7.95	1.43	47.58%	9.97%	28.57%
Max	20.00	23.00	1.11	14.00	10.00	2.00	86.77%	31.28%	58.69%
Min	14.00	16.00	0.64	9.00	6.00	1.11	25.20%	0.95%	10.88%

the time-constrained model. Table 2 shows the results for a fluctuation rate of 5%, while Table 3 shows the results for a fluctuation rate of 10%. To calculate the percent gap between the profit function of the policies and of the proposed heuristics, we use the following formula:

Percent Profit Gap
$$= \frac{\text{Profit of Proposed Heuristic - Profit of Corresponding Policy}}{\text{Profit of Proposed Heuristic}}$$
$$\times 100\% \qquad (25)$$

In Tables 2 and 3, we present the average, maximum, and minimum values obtained over all scenarios, and we observe that the allocated capacities according to the optimal policy and heuristics for office and virtual appointments fluctuate as parameters change. When the three common policies are compared, we can see that Policy-2 (i.e., $M_{s_o}/M_{s_v} = 1$) is the best, even though the results become worse as fluctuation increases. For the unconstrained model, the optimal threshold capacity allocation ratio (i.e., M_{s_o}/M_{s_v}) varies between 0.71 and 1.11 when the fluctuation rate

Table 3 Comparison of Policy-1, Policy-2, and Policy-3 with benchmark policies when the fluctuation rate is 10%

	Optimal						% Profit Gap		
Unconstrained Model	M_{s_o}	M_{s_v}	$\frac{M_{s_o}}{M_{s_v}}$	$M_{s_{v,c}}$	$M_{s_{v,u}}$	$\frac{M_{s_{v,c}}}{M_{s_{v,u}}}$	Policy-1	Policy-2	Policy-3
Average	17.41	19.74	0.89	11.61	8.12	1.47	50.69%	7.88%	31.41%
Max	22.00	25.00	1.27	17.00	12.00	2.43	106.90%	50.33%	81.14%
Min	13.00	15.00	0.62	7.00	5.00	0.88	12.34%	0.00%	3.56%
	Heuristic-1						% Profit Gap		
Model with Capacity Constraint	M_{s_o}	M_{s_v}	$\frac{M_{s_o}}{M_{s_v}}$	$M_{s_{v,c}}$	$M_{s_{v,u}}$	$\frac{M_{s_{v,c}}}{M_{s_{v,u}}}$	Policy-1	Policy-2	Policy-3
Average	15.10	16.61	0.92	9.90	6.71	1.54	47.83%	9.71%	31.07%
Max	20.00	21.00	1.50	14.00	10.00	3.33	105.82%	65.40%	81.58%
Min	10.00	12.00	0.50	6.00	3.00	0.88	3.43%	0.00%	0.00%
	Heuristic-2						% Profit Gap		
Model with Time Constraint	M_{s_o}	M_{s_v}	$\frac{M_{s_o}}{M_{s_v}}$	$M_{s_{v,c}}$	$M_{s_{v,u}}$	$\frac{M_{s_{v,c}}}{M_{s_{v,u}}}$	Policy-1	Policy-2	Policy-3
Average	16.15	19.05	0.86	11.24	7.80	1.49	52.66%	15.00%	33.45%
Max	21.00	25.00	1.29	17.00	12.00	2.60	110.39%	71.08%	96.46%
Min	13.00	14.00	0.54	7.00	5.00	0.88	11.44%	0.00%	3.08%

is 5%, while it varies between 0.62 and 1.27 when the fluctuation rate is 10%. As expected, when the uncertainty in the parameter values increases, the optimal threshold capacity allocation ratio varies more. It also shows that even if the fluctuation rate is high, it is not reasonable to use a threshold capacity allocation ratio of less than 0.62 or more than 1.27. Similar to the unconstrained model, in the model with the capacity and time constraints, Policy-2 performs the closest to the proposed solutions, but the variation is more compared to the unconstrained model where the model with time constraint has the highest variability. In the model with capacity constraint, the suggested M_{s_o}/M_{s_v} ratio varies between 0.7 and 1.14 when the fluctuation rate is 5%, and it varies between 0.5 and 1.5 when the fluctuation rate is 10%. Finally, in the time-constrained model, the proposed M_{s_o}/M_{s_v} ratio changes between 0.64 and 1.11 when the fluctuation rate is 5%, while it changes between 0.54 and 1.29 when the fluctuation rate is 10%. According to the results of the proposed policies (i.e., optimal, Heuristics 1 and 2), the average threshold capacity allocation ratio M_{s_o}/M_{s_v} should be 0.89 for the unconstrained model, 0.92 for the capacity-constrained model, and 0.86 for the time-constrained model.

We also compare the change in the total threshold capacity split for virtual appointments and for all models in Tables 2 and 3. For the unconstrained model, the optimal $M_{s_{v,c}}/M_{s_{v,u}}$ ratio varies from 1.11 to 1.86 when the fluctuation is 5%, and it varies between 0.88 and 2.43 when the fluctuation is 10%. For the model with capacity constraint, the suggested $M_{s_{v,c}}/M_{s_{v,u}}$ ratio ranges between 1.11 and 2.2 when the fluctuation rate is 5%, while this ratio ranges between 0.88 and 3.33 when the fluctuation rate is 10%. Finally, for the time-constrained model, the suggested $M_{s_{v,c}}/M_{s_{v,u}}$ ratio changes between 1.1 and 2 for the fluctuation rate of 5%, while this change is between 0.88 and 2.6 for the fluctuation rate of 10%. For all models, we observe that the average suggested $M_{s_{v,c}}/M_{s_{v,u}}$ ratio is greater than 1, which means that the threshold capacity allocated for patients diagnosed in the controlled health state at virtual appointments is greater than the threshold capacity allocated for patients who are diagnosed in the uncontrolled health state at the virtual appointments. In practice, physicians can take into account this ratio while making their recommendation to the patient for the next appointment.

We next analyze the impact of the fluctuation rate on the optimal policy for the unconstrained model. We vary the fluctuation rate between 2.5% and 20%. In Fig. 5, we present the impact of the fluctuation rate on the average, maximum, minimum, and standard deviation of the optimal M_{s_o}/M_{s_v} ratio for the unconstrained model. As the fluctuation rate increases, the average increases slightly, but the range of the ratio of office appointment threshold capacity to virtual appointment threshold capacity and the
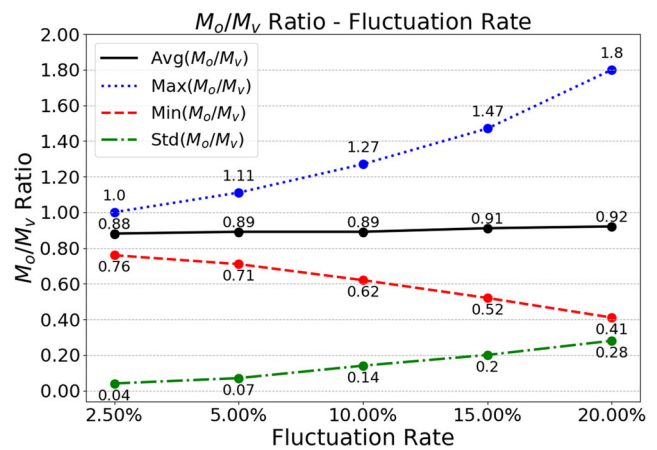
Fig. 5 The impact of fluctuation rate on the optimal threshold capacity allocation ratio, M_{s_o}/M_{s_v} for the unconstrained model

fluctuation of this ratio become larger. This indicates that the increasing uncertainty in parameter values makes the allocation decision harder for the policymakers.

6 Conclusion

Virtual appointments, consisting of e-mail, phone, and online consultations, are increasingly changing our point of view of traditional office appointments, which makes the integration of virtual appointments and office appointments critical for healthcare providers. Nowadays, many health clinics and hospitals are transitioning in virtual health services, which brings several operational challenges. Hence, the capacity planning for these two kinds of appointments has become an important but complex problem in the field of healthcare.

In this study, we use a migration network to model the flow of chronic patients. Our model further reflects the varying effectiveness of office and virtual appointments in treatment and diagnosis. We build a newsvendor optimization model to determine how to allocate the capacity of the office and virtual appointments to maximize the network's long-run average earnings. We present an unconstrained model and two extended models with capacity and time constraints to demonstrate the potential use of this model in the clinic network under different scenarios. Through numerical studies, we present one clinic network with parameters estimated from the state of Michigan. We study the use of our optimization models under different scenarios and perform sensitivity analysis for the comparison of different allocation policies. Our theoretical and numerical studies bring us several insights into the clinic system and the application of virtual appointments. We propose efficient and practical solutions that bring higher average earnings compared to the common policies in practice. Our results

also suggest that virtual appointments should be used more for following up with controlled patients than for treating uncontrolled patients. Also, although virtual appointments are not as effective as office appointments, they have equal importance to office appointments due to their lower costs.

There are several limitations that can be extended in several directions. First, our study focuses on capacity allocation decisions for given arrival and follow-up rates. As an extension, the optimal follow-up rates and arrival rates can be investigated for a given capacity of office and virtual appointments. Considering the arrival rate as a decision variable refers to the physician's panel size decision, where the physician can decide on the rate of new patients to accept into her/his panel [23, 35, 40]. Second, our data are based on the literature, which limits the generalization of the model. In the future, the model can be verified by using a specific clinic's or hospital's data. Finally, we assume in our study that all patients show up for their appointments. However, no-shows are common in practice, and no-shows in office and virtual appointments can be different. In a future study, the impact of no-shows on capacity allocation decisions can be investigated.

Appendix A: Summary of the notations

Migration Network Model

Notation	Definition
$\mu_i, \forall i \in \mathcal{I}$	The service rate at node i.
$\lambda_i, \forall i \in \mathcal{I}$	The arrival rate of new patients at node i.
$\sigma_i, \forall i \in \mathcal{I}$	Follow-up rate of patients at node i.
δ	The departure rate of patients in the system.
γ	Disease progression rate.
p_c	The probability that the patients diagnosed in the controlled health state remain in the controlled health state after a virtual appointment.
p_u	The probability that the patients diagnosed in a uncontrolled health state improve to controlled health state after a virtual appointment.
$p_{j\mid j'}, j, j' \in \mathcal{J}$	The probability that patient in health state j' is diagnosed in health state j during the virtual appointment.
p_h	The probability that the new arrived patient in virtual appointment is in the controlled health state.
w_c	System state for patients who are in the controlled health state and not scheduled for an appointment
w_u	System state for patients who are in the uncontrolled health state and not scheduled for an appointment.
s_o	System state that represents patients scheduled for office appointment.
$s_{v,c}$	System state that represents patients scheduled for virtual appointment and in the controlled health state.
$s_{v,u}$	System state that represents patients scheduled for virtual appointment and in the uncontrolled health state.
$\alpha_i, i \in \mathcal{K}'$	The expected number of patients at node i at the steady-state condition.

Capacity Allocation Model

Notation	Definition
$r_k, k \in \mathcal{K}'$	The marginal profit of appointment type k.
$\eta_k, k \in \mathcal{K}'$	The fixed cost of appointment type k.
$f_k, k \in \mathcal{K}'$	The penalty cost for overflow patients at appointment type k.
$x_k, k \in \mathcal{K}'$	The current number of patients at appointment k.
M_{s_o}	The threshold capacity of office appointments.
$M_{s_{v,c}}$	The threshold capacity of virtual appointments for controlled patients.
$M_{s_{v,u}}$	The threshold capacity of virtual appointments for uncontrolled patients.
M	Total threshold capacity for office and virtual appointments.
$A(M)$	The total long-run average earnings of the clinic.
$A_k(M_k), k \in \mathcal{K}'$	The long-run average earnings from the node i.
TC	The limited threshold capacity.

Appendix B: Proofs

1-Proof of average number of patients at each node

Recall that the model with imperfect diagnosis and treatment meets the definition of an open migration network [27, p.48], number of patients at each node satisfy the following balance equations:

$$\mu \mu_v \alpha_{s_{v,c}} - \sigma_v p_{c|c} \alpha_{w_c} - \sigma_v p_{c|u} \alpha_{w_u}$$
$$= p_h \lambda_v \tag{26}$$

$$\mu_v \alpha_{s_{v,u}} - \sigma_v (1 - p_{c|c}) \alpha_{w_c} - \sigma_v (1 - p_{c|u}) \alpha_{w_u}$$
$$= (1 - p_h) \lambda_v \tag{27}$$

$$-\mu_v p_c \alpha_{s_{v,c}} - \mu_v p_u \alpha_{s_{v,u}} + (\sigma_v + \sigma_o + \delta + \gamma) \alpha_{w_c}$$
$$-\mu_o \alpha_{s_o} = 0 \tag{28}$$

$$-\mu_v (1 - p_c) \alpha_{s_{v,c}} - \mu_v (1 - p_u) \alpha_{s_{v,u}} - \gamma \alpha_{w_c}$$
$$+(\sigma_v + \sigma_o + \delta) \alpha_{w_u} = 0 \tag{29}$$

$$-\sigma_o \alpha_{w_c} - \sigma_o \alpha_{w_u} + \mu_o \alpha_{s_o} = \lambda_o \tag{30}$$

These equations represent that the inflow to node i must be equal to outflow from node i. Equations 26–30 are five equations with five unknowns (i.e., $\alpha_{s_{v,c}}$; α_{s_o}; $\alpha_{s_{v,u}}$; α_{w_c}; α_{w_u}), then we can solve the balance equations and obtain the average number of patients in each nodes at steady state and get:

$$\alpha_{s_{v,c}} = \frac{VV_c \cdot \lambda_v + [(p_{c|u} + p_{c|c} p_u - p_{c|u} p_u) \sigma_v^2 + p_{c|c} \delta \sigma_v + p_{c|c} \sigma_o \sigma_v + p_{c|u} \gamma \sigma_v] \lambda_o}{(\delta \mu_v (\delta + \gamma + \sigma_o + \sigma_v - p_c p_{c|c} \sigma_v + p_c p_{c|u} \sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v))} \tag{31}$$

$$\alpha_{s_{v,u}} = \frac{VV_u \cdot \lambda_v + [(1 - p_{c|u} + p_c (p_{c|u} - p_{c|c})) \sigma_v^2 + (1 - p_{c|c}) \delta \sigma_v + (1 - p_{c|c}) \sigma_o \sigma_v + (1 - p_{c|u}) \gamma \sigma_v] \lambda_o}{(\delta \mu_v (\delta + \gamma + \sigma_o + \sigma_v - p_c p_{c|c} \sigma_v + p_c p_{c|u} \sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v))} \tag{32}$$

$$\alpha_{w_c} = \frac{[\sigma_o + (p_u + p_{c|u}(p_c - p_u)) \sigma_v + (p_u + p_h(p_c - p_u)) \delta] \lambda_v + [\sigma_o + (p_u + p_{c|u}(p_c - p_u)) \sigma_v + \delta] \lambda_o}{(\delta (\delta + \gamma + \sigma_o + \sigma_v - p_c p_{c|c} \sigma_v + p_c p_{c|u} \sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v))} \tag{33}$$

$$\alpha_{w_u} = \frac{[\gamma + (1 - p_u - p_{c|c}(p_c - p_u)) \sigma_v + (1 - p_u - p_h(p_c - p_u)) \delta] \lambda_v + [\gamma + (1 - p_u - p_{c|c}(p_c - p_u)) \sigma_v] \lambda_o}{(\delta (\delta + \gamma + \sigma_o + \sigma_v - p_c p_{c|c} \sigma_v + p_c p_{c|u} \sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v))} \tag{34}$$

$$\alpha_{s_o} = \frac{\sigma_o \lambda_v + (\delta + \sigma_o) \lambda_o}{\delta \mu_o} \tag{35}$$

where

$$\begin{aligned}
VV_c &= (p_{c|u} + p_{c|c} p_u - p_{c|u} p_u) \sigma_v^2 \\
&\quad + (p_{c|u} + p_h + p_{c|c} p_u - p_{c|u} p_u) \delta \sigma_v + p_{c|c} \sigma_o \sigma_v \\
&\quad + p_{c|u} \gamma \sigma_v + p_h \delta^2 + p_h \delta \gamma + p_h \delta \sigma_o \tag{36}
\end{aligned}$$

$$\begin{aligned}
VV_u &= (1 - p_{c|u} - p_c p_{c|c} + p_c p_{c|u}) \sigma_v^2 \\
&\quad + (2 - p_{c|u} - p_h - p_c p_{c|c} + p_c p_{c|u}) \delta \sigma_v \\
&\quad + (1 - p_{c|c}) \sigma_o \sigma_v + (1 - p_{c|u}) \gamma \sigma_v \\
&\quad + (1 - p_h) \delta^2 + (1 - p_h) \delta \gamma + (1 - p_h) \delta \sigma_o \tag{37}
\end{aligned}$$

From the above, we can find that,

$$\alpha_{w_c} + \alpha_{w_u} = \frac{\lambda_o + \lambda_v}{\delta} \tag{38}$$

$$\alpha_{s_{v,c}} + \alpha_{s_{v,u}} = \frac{\sigma_v \lambda_o + (\delta + \sigma_v) \lambda_v}{\delta \mu_v} \tag{39}$$

2-Proof of Theorem 1 item a

In the migration network model, the number of patients at virtual appointments (i.e., $\alpha_{s_{v,c}} + \alpha_{s_{v,u}}$) is a linearly increasing function of the virtual follow-up rate (σ_v), while it is not dependent on the office follow-up rate (σ_o).

To show the above statement we take the derivative of α_{s_v} with respect to σ_v and σ_o.

$$\frac{\partial \alpha_{s_v}}{\partial \sigma_v} = \frac{(\lambda_v + \lambda_o)}{\delta \mu_v} \tag{40}$$

$$\frac{\partial \alpha_{s_v}}{\partial \sigma_o} = 0 \tag{41}$$

As shown in Eq. 40, the first-order derivative of α_{s_v} with respect to σ_v is a positive value which states that α_{s_v} is an increasing function of σ_v. On the other hand, the first-order derivative of α_{s_v} with respect to σ_o is equal to zero which states that α_{s_v} does not depend on σ_o.

In the migration network model, the number of patients at office appointments(i.e., α_{s_o}) is an increasing function of office follow-up rate (σ_o), while it is not dependent on the virtual follow-up rate (σ_v).

To show the above statement we take the derivative of α_{s_o} with respect to σ_o and σ_v.

$$\frac{\partial \alpha_{s_o}}{\partial \sigma_o} = \frac{(\lambda_v + \lambda_o)}{\delta \mu_o} \tag{42}$$

$$\frac{\partial \alpha_{s_o}}{\partial \sigma_v} = 0 \tag{43}$$

As shown in Eq. 42, the first-order derivative of α_{s_o} with respect to σ_o is a positive value which states that α_{s_o} is an increasing function of σ_o. On the other hand, the first-order

derivative of α_{s_o} with respect to σ_v is equal to zero which states that α_{s_o} does not depend on σ_v.

3-Proof of Theorem 1 item b

In the migration network model, the number of patients who are not scheduled for an appointment and in the controlled

$$\frac{\partial \alpha_{w_c}}{\partial \sigma_o} = \frac{\left((1-p_u)(\delta\lambda_v + \lambda_o\sigma_v + \lambda_v\sigma_v) + \gamma\lambda_o + \gamma\lambda_v + (p_u - p_c)(\delta\lambda_v p_h + \lambda_o p_{c|c}\sigma_v + \lambda_v p_{c|c}\sigma_v)\right)}{(\delta(\delta + \gamma + \sigma_o + \sigma_v - p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v)^2)} \tag{44}$$

As shown with Eq. 44, the first-order derivative of α_{w_c} with respect to σ_o is a positive if value $p_u > p_c$ which means that α_{w_c} is an increasing function of σ_o. Because both

health state (i.e., α_{w_c}) is an increasing concave function of office follow-up rate (σ_o) if $p_u > p_c$.

To show the above statement we first take the derivative of α_{w_c} with respect to σ_o.

numerator and denominator yield positive values. To show that α_{w_c} is concave with respect to σ_o we take its second derivative as follows:

$$\frac{\partial^2 \alpha_{w_c}}{\partial^2 \sigma_o} = -\frac{\left(2(1-p_u)(\delta\lambda_v + \lambda_o\sigma_v + \lambda_v\sigma_v) + \gamma\lambda_o + \gamma\lambda_v + (p_u - p_c)(\delta\lambda_v p_h + \lambda_o p_{c|c}\sigma_v + \lambda_v p_{c|c}\sigma_v)\right)}{(\delta(\delta + \gamma + \sigma_o + \sigma_v - p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v)^3)} \tag{45}$$

$$= -\frac{\left(2(1-p_u)(\delta\lambda_v + \lambda_o\sigma_v + \lambda_v\sigma_v) + \gamma\lambda_o + \gamma\lambda_v + (p_u - p_c)(\delta\lambda_v p_h + \lambda_o p_{c|c}\sigma_v + \lambda_v p_{c|c}\sigma_v)\right)}{(\delta(\delta + \gamma + \sigma_o + \sigma_v + (p_u - p_c)p_{c|c}\sigma_v - (p_u - p_c)p_{c|u}\sigma_v)^3)} \tag{46}$$

As shown with Eq. 46, knowing that $p_{c|c} > p_{c|u}$, if $p_u > p_c$, the second-order derivative of α_{w_c} with respect to σ_o is a negative value which states that α_{w_c} is a concave function with respect to σ_o.

In the migration network model, the number of patients who are not scheduled for an appointment and in the

uncontrolled health state (i.e., α_{w_u}) is a decreasing convex function of office follow-up rate (σ_o) if $p_u > p_c$.

To show the above statement we first take the derivative of α_{w_c} with respect to σ_o.

$$\frac{\partial \alpha_{w_u}}{\partial \sigma_o} = \frac{(-((p_{c|c}-1)p_u - p_{c|c}p_c + 1)(\lambda_o + \lambda_v)\sigma_v - \delta\lambda_v(p_h - 1)p_u + \delta\lambda_v p_h p_c + (-\delta - \gamma)\lambda_v - \gamma\lambda_o)}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2} \tag{47}$$

$$= -\frac{(p_u - p_c)(\delta\lambda_v p_h + \lambda_o p_{c|c}\sigma_v + \lambda_v p_{c|c}\sigma_v) + (1-p_u)(\delta\lambda_v + \lambda_o\sigma_v + \lambda_v\sigma_v) + \gamma(\lambda_o + \lambda_v)}{\delta\left((p_{c|c} - p_{c|u})\sigma_v(p_u - p_c)\sigma_v + \delta + \gamma + \sigma_o + \sigma_v\right)^2} \tag{48}$$

As shown with Eq. 48, the first-order derivative of α_{w_u} with respect to σ_o is a negative value (both numerator and denominator are positive) if $p_u \geq p_c$ which states that

α_{w_u} is a decreasing function of σ_o. To show that α_{w_u} is convex with respect to σ_o we take its second derivative as follows:

$$\frac{\partial^2 \alpha_{w_u}}{\partial^2 \sigma_o} = \frac{2(p_u - p_c)(\delta\lambda_v p_h + \lambda_o p_{c|c}\sigma_v + \lambda_v p_{c|c}\sigma_v) + (1-p_u)(\delta\lambda_v + \lambda_o\sigma_v + \lambda_v\sigma_v) + \gamma(\lambda_o + \lambda_v)}{\delta\left((p_{c|c} - p_{c|u})\sigma_v(p_u - p_c)\sigma_v + \delta + \gamma + \sigma_o + \sigma_v\right)^2} \tag{49}$$

As shown with Eq. 49, the second-order derivative of α_{w_u} with respect to σ_o is a positive value (both numerator and denominator are positive) which states that α_{w_u} is a convex function with respect to σ_o.

4-Proof of Theorem 1 item c

In the migration network model, the number of patients who are not scheduled for an appointment and in the controlled

health state (i.e., α_{w_c}) is an increasing convex function of p_c, while it is an increasing concave function of p_u.

To show the above statement we first take the derivative of α_{w_c} with respect to p_c.

$$\frac{\partial \alpha_{w_c}}{\partial p_c} = \frac{\left((p_{c|c} - p_{c|u})(\delta\lambda_v p_u \sigma_v + \lambda_o p_u \sigma_v^2 + \lambda_v p_u \sigma_v^2) + \delta^2 \lambda_v p_h + \delta\gamma\lambda_v p_h + \delta\lambda_o p_{c|c}\sigma_v + \delta\lambda_v p_{c|u}\sigma_v + \delta\lambda_v p_h\sigma_o\right.}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2}$$
$$+ \frac{\left.\delta\lambda_v p_h\sigma_v + \gamma\lambda_o p_{c|u}\sigma_v + \gamma\lambda_v p_{c|u}\sigma_v + \lambda_o p_{c|c}\sigma_o\sigma_v + \lambda_o p_{c|u}\sigma_v^2 + \lambda_v p_{c|c}\sigma_o\sigma_v + \lambda_v p_{c|u}\sigma_v^2\right)}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2} \quad (50)$$

As shown with Eq. 50, knowing that $p_{c|c} > p_{c|u}$ the first-order derivative of α_{w_c} with respect to p_c is positive which means that α_{w_c} is an increasing function of p_c. To show that α_{w_c} is convex with respect to p_c we take its second derivative as follows:

$$\frac{\partial^2 \alpha_{w_c}}{\partial^2 p_c} = \frac{2\sigma_v(p_{c|c} - p_{c|u})\left((p_{c|c} - p_{c|u})(\delta\lambda_v p_u \sigma_v + \lambda_o p_u \sigma_v^2 + \lambda_v p_u \sigma_v^2) + \delta^2 \lambda_v p_h + \delta\gamma\lambda_v p_h + \delta\lambda_o p_{c|c}\sigma_v + \delta\lambda_v p_{c|u}\sigma_v\right.}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3}$$
$$+ \frac{\left.\delta\lambda_v p_h\sigma_o + \delta\lambda_v p_h\sigma_v + \gamma\lambda_o p_{c|u}\sigma_v + \gamma\lambda_v p_{c|u}\sigma_v + \lambda_o p_{c|c}\sigma_o\sigma_v + \lambda_o p_{c|u}\sigma_v^2 + \lambda_v p_{c|c}\sigma_o\sigma_v + \lambda_v p_{c|u}\sigma_v^2\right)}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3} \quad (51)$$

As shown with Eq. 51, knowing that $p_{c|c} > p_{c|u}$ the second-order derivative of α_{w_c} with respect to p_c is a positive value which states that α_{w_c} is a convex function with respect to p_c.

Next, we take the derivative of α_{w_c} with respect to p_u.

$$\frac{\partial \alpha_{w_c}}{\partial p_u} = -\frac{\left((p_{c|c} - p_{c|u})(\delta\lambda_v p_c \sigma_v + \lambda_o p_c \sigma_v^2 + \lambda_v p_c \sigma_v^2) + (p_h - 1)(\delta^2 \lambda_v + \delta\gamma\lambda_v + \delta\lambda_v\sigma_v + \delta\lambda_v\sigma_o)\right.}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2)}$$
$$+ \frac{(p_{c|c} - 1)(\delta\lambda_o\sigma_v + \lambda_o\sigma_o\sigma_v + \lambda_v\sigma_o\sigma_v) + (p_{c|u} - 1)(\delta\lambda_v\sigma_v + \gamma\lambda_o\sigma_v + \gamma\lambda_v\sigma_v + \lambda_o\sigma_v^2 + \lambda_v\sigma_v^2)\right)}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2)} \quad (52)$$

As shown with Eq. 52, the first-order derivative of α_{w_c} with respect to p_u is positive (since $(\delta\lambda_v\sigma_v + \gamma\lambda_o\sigma_v + \gamma\lambda_v\sigma_v + \lambda_o\sigma_v^2 + \lambda_v\sigma_v^2) \geq (p_{c|c}(\delta\lambda_v p_c\sigma_v + \lambda_o p_c\sigma_v^2 + $ $\lambda_v p_c\sigma_v^2))$ which means that α_{w_c} is an increasing function of p_u. To show that α_{w_c} is concave with respect to p_u we take its second derivative as follows:

$$\frac{\partial^2 \alpha_{w_c}}{\partial^2 p_u} = \frac{2\sigma_v(p_{c|c} - p_{c|u})\left((p_{c|c} - p_{c|u})(\delta\lambda_v p_c \sigma_v + \lambda_o p_c \sigma_v^2 + \lambda_v p_c \sigma_v^2) + (p_h - 1)(\delta^2\lambda_v + \delta\gamma\lambda_v + \delta\lambda_v\sigma_v + \delta\lambda_v\sigma_o)\right.}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3)}$$
$$+ \frac{(p_{c|c} - 1)(\delta\lambda_o\sigma_v + \lambda_o\sigma_o\sigma_v + \lambda_v\sigma_o\sigma_v) + (p_{c|u} - 1)(\delta\lambda_v\sigma_v + \gamma\lambda_o\sigma_v + \gamma\lambda_v\sigma_v + \lambda_o\sigma_v^2 + \lambda_v\sigma_v^2)\right)}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3)} \quad (53)$$

As shown with Eq. 53, knowing that $p_{c|c} > p_{c|u}$, the second-order derivative of α_{w_c} with respect to p_c is a negative value which states that α_{w_c} is a concave function with respect to p_u.

In the migration network model, the number of patients who are not scheduled for an appointment and in the uncontrolled health state (i.e., α_{w_u}) is a decreasing concave function of p_c, while it is a decreasing convex function of p_u.

To show the above statement we first take the derivative of α_{w_u} with respect to p_c.

$$
\begin{aligned}
\frac{\partial \alpha_{w_u}}{\partial p_c} = &-\frac{\big((p_{c|c} - p_{c|u})(\delta\lambda_v p_u\sigma_v + \lambda_o p_u\sigma_v^2 + \lambda_v p_u\sigma_v^2) + \delta^2\lambda_v p_h + \delta\gamma\lambda_v p_h + \delta\lambda_o p_{c|c}\sigma_v + \delta\lambda_v p_{c|u}\sigma_v + \delta\lambda_v p_h\sigma_o}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u\sigma_v - p_{c|u} p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2} \\
&+\frac{\delta\lambda_v p_h\sigma_v + \gamma\lambda_o p_{c|c}\sigma_v + \gamma\lambda_v p_{c|u}\sigma_v + \lambda_o p_{c|c}\sigma_o\sigma_v + \lambda_o p_{c|u}\sigma_v^2 + \lambda_v p_{c|c}\sigma_o\sigma_v + \lambda_v p_{c|u}\sigma_v^2\big)}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u\sigma_v - p_{c|u} p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2}
\end{aligned}
\tag{54}
$$

As shown with Eq. 54, knowing that $p_{c|c} > p_{c|u}$ the first-order derivative of α_{w_u} with respect to p_c is negative which means that α_{w_u} is a decreasing function of p_c. To show that α_{w_u} is concave with respect to p_c we take its second derivative as follows:

$$
\begin{aligned}
\frac{\partial^2 \alpha_{w_u}}{\partial^2 p_c} = &-\frac{2\sigma_v(p_{c|c} - p_{c|u})\big((p_{c|c} - p_{c|u})(\delta\lambda_v p_u\sigma_v + \lambda_o p_u\sigma_v^2 + \lambda_v p_u\sigma_v^2) + \delta^2\lambda_v p_h + \delta\gamma\lambda_v p_h + \delta\lambda_o p_{c|c}\sigma_v}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u\sigma_v - p_{c|u} p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3} \\
&+\frac{\delta\lambda_v p_{c|u}\sigma_v + \delta\lambda_v p_h(\sigma_o + \sigma_v) + \gamma p_{c|u}\sigma_v(\lambda_o + \lambda_v) + \lambda_o p_{c|c}\sigma_o\sigma_v + \lambda_o p_{c|u}\sigma_v^2 + \lambda_v p_{c|c}\sigma_o\sigma_v + \lambda_v p_{c|u}\sigma_v^2}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u\sigma_v - p_{c|u} p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3}
\end{aligned}
\tag{55}
$$

As shown with Eq. 55, knowing that $p_{c|c} > p_{c|u}$ the second-order derivative of α_{w_u} with respect to p_c is a negative value which states that α_{w_u} is a concave function with respect to p_c.

Next, we take the derivative of α_{w_u} with respect to p_u.

$$
\begin{aligned}
\frac{\partial \alpha_{w_u}}{\partial p_u} = &\frac{\big((p_{c|c} - p_{c|u})(\delta\lambda_v p_c\sigma_v + \lambda_o p_c\sigma_v^2 + \lambda_v p_c\sigma_v^2) + (p_h - 1)(\delta^2\lambda_v + \delta\gamma\lambda_v + \delta\lambda_v\sigma_v + \delta\lambda_v\sigma_o)}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u\sigma_v - p_{c|u} p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2)} \\
&+\frac{(p_{c|c} - 1)(\delta\lambda_o\sigma_v + \lambda_o\sigma_o\sigma_v + \lambda_v\sigma_o\sigma_v) + (p_{c|u} - 1)(\delta\lambda_v\sigma_v + \gamma\lambda_o\sigma_v + \gamma\lambda_v\sigma_v + \lambda_o\sigma_v^2 + \lambda_v\sigma_v^2)\big)}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u\sigma_v - p_{c|u} p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2)}
\end{aligned}
\tag{56}
$$

As shown with Eq. 56, the first-order derivative of α_{w_u} with respect to p_u is negative (since $(\delta\lambda_v\sigma_v + \gamma\lambda_o\sigma_v + \gamma\lambda_v\sigma_v + \lambda_o\sigma_v^2 + \lambda_v\sigma_v^2) \geq (p_{c|c}(\delta\lambda_v p_c\sigma_v + \lambda_o p_c\sigma_v^2 + \lambda_v p_c\sigma_v^2))$ which means that α_{w_u} is a decreasing function of p_u. To show that α_{w_u} is convex with respect to p_u we take its second derivative as follows:

$$
\begin{aligned}
\frac{\partial^2 \alpha_{w_u}}{\partial^2 p_u} = &-\frac{2\sigma_v(p_{c|c} - p_{c|u})\big((p_{c|c} - p_{c|u})(\delta\lambda_v p_c\sigma_v + \lambda_o p_c\sigma_v^2 + \lambda_v p_c\sigma_v^2) + (p_h - 1)(\delta^2\lambda_v + \delta\gamma\lambda_v + \delta\lambda_v\sigma_v + \delta\lambda_v\sigma_o)}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u\sigma_v - p_{c|u} p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3)} \\
&+\frac{(p_{c|c} - 1)(\delta\lambda_o\sigma_v + \lambda_o\sigma_o\sigma_v + \lambda_v\sigma_o\sigma_v) + (p_{c|u} - 1)(\delta\lambda_v\sigma_v + \gamma\lambda_o\sigma_v + \gamma\lambda_v\sigma_v + \lambda_o\sigma_v^2 + \lambda_v\sigma_v^2)\big)}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u\sigma_v - p_{c|u} p_u\sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3)}
\end{aligned}
\tag{57}
$$

As shown with Eq. 57, knowing that $p_{c|c} > p_{c|u}$, the second-order derivative of α_{w_u} with respect to p_u is a positive value which states that α_{w_u} is a convex function with respect to p_u.

5-Proof of Theorem 1 item d

In the migration network model, the number of patients who are not scheduled for an appointment and in the controlled

health state (i.e., α_{w_c}) is a decreasing convex function of $p_{c|c}$, while it is a decreasing concave function of $p_{c|u}$ if $p_u > p_c$.

To show the above statement we first take the derivative of α_{w_c} with respect to $p_{c|c}$.

$$\frac{\partial \alpha_{w_c}}{\partial p_{c|c}} = \frac{\sigma_v(p_c - p_u)\Big((1 - p_h)\delta\lambda_v p_u + (1 - p_{c|u})\lambda_o p_u \sigma_v + (1 - p_{c|u})\lambda_v p_u \sigma_v + \delta\lambda_v p_c p_h + \lambda_o p_c p_{c|u}\sigma_v}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2)}$$
$$+ \frac{\lambda_v p_c p_{c|u}\sigma_v + \delta\lambda_o + \lambda_o\sigma_o + \lambda_v\sigma_o\Big)}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2)} \tag{58}$$

As shown with Eq. 58, knowing that $p_{c|c} > p_{c|u}$, if $p_u > p_c$ the first-order derivative of α_{w_c} with respect to $p_{c|c}$ is negative which means that α_{w_c} is a decreasing function

of $p_{c|c}$. To show that α_{w_c} is convex with respect to $p_{c|c}$ we take its second derivative as follows:

$$\frac{\partial^2 \alpha_{w_c}}{\partial^2 p_{c|c}} = \frac{2\sigma_v^2(p_c - p_u)^2\Big((1 - p_h)\delta\lambda_v p_u + (1 - p_{c|u})\lambda_o p_u \sigma_v + (1 - p_{c|u})\lambda_v p_u \sigma_v + \delta\lambda_v p_c p_h + \lambda_o p_c p_{c|u}\sigma_v}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3)}$$
$$+ \frac{\lambda_v p_c p_{c|u}\sigma_v + \delta\lambda_o + \lambda_o\sigma_o + \lambda_v\sigma_o\Big)}{(\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3)} \tag{59}$$

As shown with Eq. 59, knowing that $p_{c|c} > p_{c|u}$ the second-order derivative of α_{w_c} with respect to $p_{c|c}$ is a

positive value which states that α_{w_c} is a convex function with respect to $p_{c|c}$.

Next, we take the derivative of α_{w_c} with respect to $p_{c|u}$.

$$\frac{\partial \alpha_{w_c}}{\partial p_{c|u}} = -\frac{(p_{c|c}(\lambda_o + \lambda_v)\sigma_v + \delta\lambda_v p_h)(p_c - p_u)^2\sigma_v + (p_c - p_u)\sigma_v\Big(((p_u - 1)\sigma_v}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2}$$
$$+ \frac{p_u\delta - \delta - \gamma)\lambda_v + ((p_u - 1)\sigma_v - \gamma)\lambda_o\Big)}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^2} \tag{60}$$

As shown with Eq. 60, the first-order derivative of α_{w_c} with respect to $p_{c|u}$ is negative if $p_u > p_c$ which means that α_{w_c} is a decreasing function of $p_{c|u}$. To show that α_{w_c}

is concave with respect to $p_{c|u}$ we take its second derivative as follows:

$$\frac{\partial^2 \alpha_{w_c}}{\partial^2 p_{c|u}} = \frac{2(p_c\sigma_v - p_u\sigma_v)\Big((p_{c|c}(\lambda_o + \lambda_v)\sigma_v + \delta\lambda_v p_h)(p_c - p_u)^2\sigma_v + (p_c - p_u)\sigma_v\Big(((p_u - 1)\sigma_v}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3}$$
$$+ \frac{p_u\delta - \delta - \gamma)\lambda_v + ((p_u - 1)\sigma_v - \gamma)\lambda_o\Big)\Big)}{\delta(-p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c} p_u \sigma_v - p_{c|u} p_u \sigma_v + \delta + \gamma + \sigma_o + \sigma_v)^3} \tag{61}$$

As shown with Eq. 61, knowing that $p_{c|c} > p_{c|u}$, if $p_u > p_c$ the second-order derivative of α_{w_c} with respect to p_c is a negative value which states that α_{w_c} is a concave function with respect to $p_{c|u}$.

In the migration network model, the number of patients who are not scheduled for an appointment and in the

uncontrolled health state (i.e., α_{w_u}) is an increasing convex function of $p_{c|u}$ if $p_u > p_c$.

To show the above statement we first take the derivative of α_{w_u} with respect to $p_{c|u}$.

$$\frac{\partial \alpha_{w_u}}{\partial p_{c|u}} = \frac{\sigma_v(p_c - p_u)\Big(((-p_{c|c} + 1)p_u + p_c p_{c|c} - 1)(\lambda_o + \lambda_v)\sigma_v + (-\delta(ph - 1)p_u + \delta p_c ph - \delta - \gamma)\lambda_v - \gamma\lambda_o\Big)}{((((-p_{c|c} + p_{c|u})p_u + p_c p_{c|c} - p_c p_{c|u} - 1)\sigma_v - \delta - \gamma - \sigma_o)^2 \delta)} \quad (62)$$

As shown with Eq. 62, if $p_u > p_c$ the first-order derivative of α_{w_c} with respect to $p_{c|c}$ is positive since both the numerator and denominator is positive. This means that

α_{w_u} is an increasing function of $p_{c|u}$. To show that α_{w_u} is convex with respect to $p_{c|u}$ we take its second derivative as follows:

$$\frac{\partial^2 \alpha_{w_u}}{\partial^2 p_{c|u}} = \frac{2\sigma_v^2(p_c - p_u)^2\Big((((-p_{c|c} + 1)p_u + p_c p_{c|c} - 1)(\lambda_o + \lambda_v)\sigma_v + (-\delta(p_h - 1)p_u + \delta p_c p_h - \delta - \gamma)\lambda_v - \gamma\lambda_o\Big)}{((((-p_{c|c} + p_{c|u})p_u + p_c p_{c|c} - p_c p_{c|u} - 1)\sigma_v - \delta - \gamma - \sigma_o)^3 \delta)}$$

$$(63)$$

As shown with Eq. 63, knowing that $p_{c|c} > p_{c|u}$, if $p_u > p_c$, the second-order derivative of α_{w_u} with respect to $p_{c|u}$ is a positive value which states that α_{w_c} is a convex function with respect to $p_{c|c}$. More specifically, the numerator is negative and the denominator is also negative hence the overall calculation is a positive value.

6-Proof of Theorem 1 item e

In the migration network model, the number of patients who are not scheduled for an appointment and in the both controlled and uncontrolled health States (i.e., α_{w_c}, and α_{w_u}) are not dependent on the service rate of office and virtual appointments (μ_o and μ_v).

To show the above statement, we first take the derivative of α_{w_c} with respect to μ_o and μ_v as follows:

$$\alpha_{w_c} = \frac{[\sigma_o + (p_u + p_{c|u}(p_c - p_u))\sigma_v + (p_u + p_h(p_c - p_u))\delta]\lambda_v + [\sigma_o + (p_u + p_{c|u}(p_c - p_u))\sigma_v + \delta]\lambda_o}{(\delta(\delta + \gamma + \sigma_o + \sigma_v - p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v))} \quad (64)$$

$$\frac{\partial \alpha_{w_c}}{\partial \mu_o} = \frac{\partial \alpha_{w_c}}{\partial \mu_v} = 0 \quad (65)$$

Similarly, the term α_{w_u} and its first order derivative with respect to μ_o and μ_v can be defined as follows:

$$\alpha_{w_u} = \frac{[\gamma + (1 - p_u - p_{c|c}(p_c - p_u))\sigma_v + (1 - p_u - p_h(p_c - p_u))\delta]\lambda_v + [\gamma + (1 - p_u - p_{c|c}(p_c - p_u))\sigma_v]\lambda_o}{(\delta(\delta + \gamma + \sigma_o + \sigma_v - p_c p_{c|c}\sigma_v + p_c p_{c|u}\sigma_v + p_{c|c}p_u\sigma_v - p_{c|u}p_u\sigma_v))} \quad (66)$$

$$\frac{\partial \alpha_{w_u}}{\partial \mu_o} = \frac{\partial \alpha_{w_u}}{\partial \mu_v} = 0 \quad (67)$$

Since the derivatives of both α_{w_c} and α_{w_u} with respect to μ_o and μ_v are equal to zero, they are not dependent on the office and virtual service rates.

7-Proof of reformulation of the objective function

The original objective function of the unconstrained capacity allocation model is:

$$A(M) = \lim_{T \to \infty} \frac{1}{T} \left\{ \sum_{k \in \mathcal{K}'} \int_0^T r_k min\, [x_k(t), M_k]\, dt \right. \tag{68}$$
$$\left. - \sum_{k \in \mathcal{K}'} \int_0^T \eta_k M_k dt - \sum_{k \in \mathcal{K}'} \int_0^T f_k\, (x_k(t) - M_k)^+\, dt \right\}$$

To simplify the objective function (68), we make some changes:

$$min\, [x_k(t), M_k] = x_k(t) - [x_k(t) - M_k]^+ \tag{69}$$
$$M_k = x_k(t) - [x_k(t) - M_k]^+ + [M_k - x_k(t)]^+ \tag{70}$$

Substitute Eqs. 69 and 70 into objective function (68) gives:

$$A(M) = \sum_{k \in \mathcal{K}'} \lim_{T \to \infty} \frac{1}{T} \left\{ \int_0^T (r_k - \eta_k) x_k(t) dt \right.$$
$$- \int_0^T \eta_k\, [M_k - x_k(t)]^+\, dt$$
$$\left. - \int_0^T (f_k + r_k - \eta_k)\, [x_k(t) - M_k]^+\, dt \right\} \tag{71}$$

Due to ergodicity, let $\mathbb{E}_{\pi_k}(x_k), k \in \mathcal{K}'$ be the expected number of patients at node k under the steady-state distribution $\pi_k, k \in \mathcal{K}'$. We have:

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T x_k(t) dt = \mathbb{E}_{\pi_k}(x_k) = \alpha_k \tag{72}$$

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T [M_k - x_k(t)]^+\, dt = \mathbb{E}_{\pi_k}\big[(M_k - x_k)^+\big] \tag{73}$$

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T [x_k(t) - M_k]^+\, dt = \mathbb{E}_{\pi_k}\big[(x_k - M_k)^+\big] \tag{74}$$

Take Eqs. 72–74 into objective function (71), we get:

$$A(M) = \sum_{k \in \mathcal{K}'} \Big[(r_k - \eta_k)\alpha_k - \eta_k \mathbb{E}_{\pi_k}\big[(M_k - x_k)^+\big]$$
$$- (f_k + r_k - \eta_k)\mathbb{E}_{\pi_k}\big[(x_k - M_k)^+\big] \Big] \tag{75}$$

In the Function (75), the first term is the marginal profit. The second term is the opportunity cost for unutilized capacity. The last term represents the cost due to patient overflow.

8-Proof of Proposition 1

Since the office and virtual processes are independent to each other, to maximize the objective $A(M)$, it suffices to maximize the sub-objective $A_k(M_k)$ separately.

$$A_k(M_k) = (r_k - \eta_k)\alpha_k - \eta_k \mathbb{E}_{\pi_k}\big[(M_k - x_k)^+\big]$$
$$- (f_k + r_k - \eta_k)\mathbb{E}_{\pi_k}\big[(x_k - M_k)^+\big], \forall k \in \mathcal{K}' \tag{76}$$

We note that $(M_k - x_k)^+ = \max\{M_k - x_k, 0\}$ so that the difference yields a positive value. Then, the term $\eta_k \mathbb{E}_{\pi_k}\big[(M_k - x_k)^+\big]$ can be defined as follows: $\eta_k \mathbb{E}_{\pi_k} \big[(M_k - x_k)^+\big] = \eta_k \sum_{x_k=0}^{M_k}(M_k - x_k)^+\pi_k(x_k) + \eta_k \sum_{x_k=M_k}^{\infty}(M_k - x_k)^+\pi_k(x_k) = \eta_k \sum_{x_k=0}^{M_k}(M_k - x_k)\pi_k(x_k) + \eta_k \sum_{x_k=M_k}^{\infty} 0\pi_k(x_k) = \eta_k \sum_{x_k=0}^{M_k}(M_k - x_k)\pi_k(x_k) + 0$. The term $(f_k + r_k - \eta_k)\mathbb{E}_{\pi_k}\big[(x_k - M_k)^+\big]$ can be expressed in a similar way as well.

Hence, the objective function $A_k(M_k)$ can be defined as follows:

$$A_k(M_k) = (r_k - \eta_k)\alpha_k - \eta_k \mathbb{E}_{\pi_k}\big[(M_k - x_k)^+\big]$$
$$- (f_k + r_k - \eta_k)\mathbb{E}_{\pi_k}\big[(x_k - M_k)^+\big], \forall k \in \mathcal{K}' \tag{77}$$
$$= (r_k - \eta_k)\alpha_k - \eta_k \sum_{x_k=0}^{M_k}(M_k - x_k)\pi_k(x_k)$$
$$- (f_k + r_k - \eta_k) \sum_{x_k=M_k}^{\infty}(x_k - M_k)\pi_k(x_k),$$
$$\forall k \in \mathcal{K}' \tag{78}$$

Similarly, $A_k(M_k + 1)$ is defined as follows:

$$A_k(M_k + 1) = (r_k - \eta_k)\alpha_k - \eta_k \mathbb{E}_{\pi_k}(M_k + 1 - x_k)^+$$
$$- (f_k + r_k - \eta_k)\mathbb{E}_{\pi_k}(x_k - (M_k + 1))^+, \forall k \in \mathcal{K}' \tag{79}$$
$$= (r_k - \eta_k)\alpha_k - \eta_k \sum_{x_k=0}^{M_k+1}(M_k + 1 - x_k)\pi_k(x_k)$$
$$- (f_k + r_k - \eta_k) \sum_{x_k=M_k+1}^{\infty}(x_k - (M_k + 1))\pi_k(x_k),$$
$$\forall k \in \mathcal{K}' \tag{80}$$

Then, the differential and the second-order differential of function (76) are as follows:

$$\frac{\Delta A_k(M_k)}{\Delta(M_k)} = A_k(M_k + 1) - A_k(M_k) \tag{81}$$

$$\left[(r_k - \eta_k)\alpha_k - \eta_k \sum_{x_k=0}^{M_k} (M_k - x_k)\pi_k(x_k) \right.$$

$$\left. - (f_k + r_k - \eta_k) \sum_{x_k=M_k}^{\infty} (x_k - M_k)\pi_k(x_k) \right]$$

$$- \left[(r_k - \eta_k)\alpha_k - \eta_k \sum_{x_k=0}^{M_k+1} (M_k + 1 - x_k)\pi_k(x_k) \right.$$

$$\left. - (f_k + r_k - \eta_k) \sum_{x_k=M_k+1}^{\infty} (x_k - (M_k + 1))\pi_k(x_k) \right] \tag{82}$$

$$= -\eta_k \left[\sum_{x_k=0}^{M_k+1} (M_k+1-x_k)\pi_k(x_k) - \sum_{x_k=0}^{M_k} (M_k-x_k)\pi_k(x_k) \right]$$

$$- (f_k + r_k - \eta_k) \left[\sum_{x_k=M_k+1}^{\infty} (x_k - (M_k + 1))\, \pi_k(x_k) \right.$$

$$\left. - \sum_{x_k=M_k}^{\infty} (x_k - M_k)\pi_k(x_k) \right] \tag{83}$$

$$= -\eta_k \sum_{x_k=0}^{M_k} \pi_k(x_k) + (f_k + r_k - \eta_k) \sum_{x_k=M_k+1}^{\infty} \pi_k(x_k) \tag{84}$$

$$= -\eta_k \pi_k(x_k \le M_k) + (f_k + r_k - \eta_k)\pi_k(x_k > M_k) \tag{85}$$

$$= f_k + r_k - \eta_k - (f_k + r_k)\pi_k(x_k \le M_k) \tag{86}$$

$$\frac{\Delta^2 A_k(M_k)}{\Delta^2(M_k)} = \Delta A_k(M_k + 1) - \Delta A_k(M_k)$$
$$= -(f_k + r_k)\left[\pi_k(x_k \le M_k + 1) \right.$$
$$\left. - \pi_k(x_k \le M_k) \right] < 0 \tag{87}$$

It is clear that the objective function $A_k(M_k)$ is a discrete concave function. Hence, to maximize $A_k(M_k)$, the optimal threshold capacity of node k is the smallest positive integer $M_k = M_k^{min}$ that makes $\Delta A_k(M_k) \le 0$, and we have

$$\forall M_k \in [0, M_k^{min}), \ \Delta A_k M_k > 0, \ then, \ A_k(M_k)$$
$$< A_k(M_k^{min}) \tag{88}$$
$$\forall M_k \in [M_k^{min}, \infty), \ \Delta A_k M_k \le 0, \ then, \ A_k(M_k)$$
$$\le A_k(M_k^{min}) \tag{89}$$

In one case that if $\Delta A_k(M_k^{min}) = 0$, the optimal threshold capacity can be M_k^{min} or $(M_k^{min} + 1)$, since $A_k(M_k^{min}) = A_k(M_k^{min} + 1)$. However, it has a small probability that $\Delta A_k(M_k^{min}) = 0$ and even in that case, M_k^{min} is one of the optimal solutions. Hence, we conclude that M_k^{min} is the optimal threshold capacity for node k that maximize $A_k(M_k)$.

$$M_k^{min} = min \left\{ M_k \in N_+ : \pi_k(x_k \le M_k) \ge \frac{f_k + r_k - \eta_k}{f_k + r_k} \right\} \tag{90}$$

From Function (90), we obtain $M = (M_{s_o}^{min}, M_{s_{v,c}}^{min}, M_{s_{v,u}}^{min})$, which is the optimal threshold capacity for the $A(M)$.

9-Proof of Proposition 2

When physicians do not have enough working time, the solution is obtained through the Heuristic 2. Assume through t^{th} iteration, we obtain the solution from the heuristic, and the solution is $M_{T_w} = M_{T_w}^t = (M_{s_o}{}^t, M_{s_{v,c}}^t, M_{s_{v,u}}^t)$. Hence, M_{T_w} satisfies:

$$\sum_{k \in \mathcal{K}'} \frac{1}{\mu_k} M_k^t \le T_w \tag{91}$$

$$\sum_{k \in \mathcal{K}'} \frac{1}{\mu_k} M_k^t + \frac{1}{\mu_{k'}} > T_w \quad \forall k' \in \mathcal{K}' \tag{92}$$

Now, if we relax the time constraint and let the heuristic runs one more iteration, we have $M_{T_w}^{t+1} = M_{T_w} + e_x$, where e_x is the xth unit vector, and $x = argmax_{k'} \ \mu_{k'} A'(M_{k'}^t)$. Refer to [21], we have the following equations for Heuristic 1:

$$A(M_{TC}^{t+1}) > A(M_{TC}^*) \ge A(M_{TC}) \tag{93}$$

$$\sum_{k \in \mathcal{K}'} M_k^{t+1} > TC \ge \sum_{k \in \mathcal{K}'} M_{k,TC}^* \ge \sum_{k \in \mathcal{K}'} M_k^t \tag{94}$$

and the following equations for Heuristic 2:

$$A(M_{T_w}^{t+1}) > A(M_{T_w}^*) \ge A(M_{T_w}) \tag{95}$$

$$\sum_{k \in \mathcal{K}'} \frac{1}{\mu_k} M_k^{t+1} > T_w \ge \sum_{k \in \mathcal{K}'} \frac{1}{\mu_k} M_{k,T_w}^* \ge \sum_{k \in \mathcal{K}'} \frac{1}{\mu_k} M_k^t \tag{96}$$

Inequality Eqs. 93 and 94 show that the optimal average long-run earnings and the threshold capacity are between those under the sub-optimal solution M_{TC} and the solution M_{TC}^{t+1} that we allow to run one more iteration, respectively. Similarly, Inequality Eqs. 95 and 96 show that the optimal average long-run earnings and the working time are between those under the sub-optimal solution M_{T_w} and the solution

$M_{T_w}^{t+1}$ that we allow to run one more iteration. Then By considering inequality (93), we have the following equation:

$$A(M_{TC}^*) - A(M_{TC}) < A(M_{TC}^{t+1}) - A(M_{TC}) = A'(M_x^t)$$
$$\leq max \ A'(M_k^t) \quad (97)$$

where $x = argmax_{k'} A'(M_{k'}^t)$. Similarly, by considering inequality (95), we have the following equation:

$$A(M_{T_w}^*) - A(M_{T_w}) < A(M_{T_w}^{t+1}) - A(M_{T_w}) = A'(M_x^t)$$
$$\leq max \ A'(M_k^t) \quad (98)$$

where $x = argmax_{k'} \ \mu_{k'} A'(M_{k'}^t)$.

With inequality Eqs. 93 and 97, we can show that the relative error by using the solution from Heuristic 1, M_{TC}, as an approximation of M_{TC}^* is no greater than $\frac{max \ A'(M_k^t)}{A(M_{TC})}$.

$$\frac{A(M_{TC}^*) - A(M_{TC})}{A(M_{TC}^*)} \leq \frac{A(M_{TC}^*) - A(M_{TC})}{A(M_{TC})}$$
$$< \frac{max_k \ A'(M_k^t)}{A(M_{TC})} \quad (99)$$

Similarly, with inequality Eqs. 95 and 98, we can show that the relative error by using the solution from Heuristic 2, M_{T_w}, as an approximation of $M_{T_w}^*$ is no greater than $\frac{max \ A'(M_k^t)}{A(M_{T_w})}$.

$$\frac{A(M_{T_w}^*) - A(M_{T_w})}{A(M_{T_w}^*)} \leq \frac{A(M_{T_w}^*) - A(M_{T_w})}{A(M_{T_w})}$$
$$< \frac{max \ A'(M_k^t)}{A(M_{T_w})} \quad (100)$$

References

1. Ackerman MJ, Filart R, Burgess LP, Lee I, Poropatich RK (2010) Developing next-generation telehealth tools and technologies: patients, systems, and data perspectives. Telemedicine and e-Health 16(1):93–95
2. American Hospital Association (2016) Telehealth: helping hospitals deliver cost-effective care. Available at: https://www.aha.org/system/files/content/16/16telehealthissuebrief.pdf. Accessed June 2020
3. American Telemedicine Association (2020) What is telemedicine& telehealth? Available at https://s3.amazonaws.com/rdcms-himss/files/production/public/HIMSSorg/Content/files/Line%2016%20%20What%20Is%20Telemedicine.pdf. Accessed June 2020
4. Bashshur RL (1995) On the definition and evaluation of telemedicine. Telemedicine Journal 1(1):19–30
5. Bavafa H, Hitt LM, Terwiesch C (2018) The impact of e-visits on visit frequencies and patient health: evidence from primary care. Manag Sci 64(12):5461–5480
6. Bavafa H, Savin S, Terwiesch C (2019) Redesigning primary care delivery: customized office revisit intervals and e-visits. Available at https://papers.ssrn.com/sol3/papers.cfm?abstractid=2363685. Accessed June 2020
7. Bayram A, Deo S, Iravani S, Smilowitz K (2019) Managing virtual appointments in chronic care. IISE Transactions on Healthcare Systems Engineering 10(1):1–17
8. Bedi BS (2009) Telemedicine standards: need and Indian initiatives. Telemedicine and e-Health 15(6):597–599
9. Block HW, Langberg NA, Savits TH (1993) Repair replacement policies. J Appl Probab 30(1):194–206
10. Bodenheimer T, Wagner EH, Grumbach K (2002) Improving primary care for patients with chronic illness. Jama 288(14):1775–1779
11. Bretthauer KM, Cote MJ (1998) A model for planning resource requirements in health care organizations. Decis Sci 29(1):243–270
12. Caceres C, Gomez EJ, Garcia F, Gatell JM, del Pozo F (2006) An integral care telemedicine system for HIV/AIDS patients. Int J Med Inform 75(9):638–642
13. Behavioral risk factor surveillance system survey (BRFSS), Atlanta, Georgia: US Department of Health and Human Services (2010) Centers for Disease Control and Prevention. Available at: https://www.cdc.gov/brfss/index.html. Accessed June 2020
14. Centers for Disease Control and Prevention (2009). The Power of Prevention. Available at: https://www.cdc.gov/chronicdisease/pdf/2009-Power-of-Prevention.pdf. Accessed June 2020
15. Clinics.com (2019). Clinics in Michigan. Available at: https://www.freeclinics.com/sta/michigan. Accessed June 2020
16. Craig J, Chua R, Russell C, Patterson V, Wootton R (2000) The cost-effectiveness of teleneurology consultations for patients admitted to hospitals without neurologists on site. 1: a retrospective comparison of the case-mix and management at two rural hospitals. J Telemed Telecare 6(1 suppl):46–49
17. Dall TM, Roary M, Yang W, Zhang S, Zhang Y, Arday DR, Gantt CJ, Chen YJ (2011) Peer-reviewed: health care use and costs for participants in a diabetes disease management program, USA 2007–2008. Preventing Chronic Disease 8(3):A53. Published online. PMCID:PMC3103558
18. Deo S, Iravani S, Jiang T, Smilowitz K, Samuelson S (2013) Improving health outcomes through better capacity allocation in a community-based chronic care model. Oper Res 61(6):1277–1294
19. Department of Health and Human Services (2016). Chronic care management services. ICN December
20. Felzen M, Beckers SK, Kork F, Hirsch F, Bergrath S, Sommer A, Brokmann JC, Czaplik M, Rossaint R (2019) Utilization, safety, and technical performance of a telemedicine system for prehospital emergency care: observational study. Journal of Medical Internet Research 21(10). https://doi.org/10.2196/14907
21. Fox B (1966) Discrete optimization via marginal analysis. Manag Sci 13(3):210–216
22. Goodarzi M, Ebrahimzadeh I, Rabi A, Saedipoor B, Jafarabadi MA (2012) Impact of distance education via mobile phone text messaging on knowledge, attitude, practice and self-efficacy of patients with type 2 diabetes mellitus in Iran. Journal of Diabetes & Metabolic Disorders 11(1):1–10
23. Green LV, Savin S (2008) Reducing delays for medical appointments: a queueing approach. Oper Res 56(6):1526–1538
24. Gupta D, Denton B (2008) Appointment scheduling in health care: challenges and opportunities. IIE Trans 40(9):800–819
25. Heffler S, Smith S, Won G, Clemens MK, Keehan S, Zezza M (2002) Health spending projections for 2001–2011: the latest outlook. Health Aff 21(2):207–218
26. Hulshof PJ, Kortbeek N, Boucherie RJ, Hans EW, Bakker PJ (2012) Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. Health Systems 1(2):129–175
27. Kelly FP (1979) Reversibility and stochastic networks. Cambridge University Press

28. Khosla S (2020) Implementation of synchronous telemedicine into clinical practice. Sleep Medicine Clinics 15(3):347

29. Kochanek KD, Minino AM, Murphy SL, Xu J, Kung H-C (2011) Deaths: final data for 2009. National Vital Statistics Reports

30. Kucukyazici B, Verter V, Mayao NE (2011) An analytical framework for designing community-based care for chronic diseases. Prod Oper Manag 20(3):474–488

31. Kucukyazici B, Verter V (2013) Managing community-based care for chronic diseases: the quantitative approach. Operations Research and Health Care Policy, Springer, pp 71–90

32. Lee DK, Zenios SA (2009) Optimal capacity overbooking for the regular treatment of chronic conditions. Oper Res 57(4):852–865

33. Leggett P, Graham L, Steele K, Gilliland A, Stevenson M, O'Reilly D, Wootton R, Taggart A (2001) Telerheumatology– diagnostic accuracy and acceptability to patient, specialist, and general practitioner. Br J Gen Pract 51(470):746–748

34. Li Y, Zhang Y, Kong N, Lawley M (2016) Capacity planning for long-term care networks. IIE Trans 48(12):1098–1111

35. Liu N, Ziya S (2014) Panel size and overbooking decisions for appointment-based services under patient no-shows. Production and Operations Management 23(12):2209002223

36. Liu X, Hu M, Helm JE, Lavieri MS, Skolarus TA (2018) Missed opportunities in preventing hospital readmissions: redesigning post-discharge checkup policies. Prod Oper Manag 27(12):2226–2250

37. Marcin JP (2013) Telemedicine in the pediatric intensive care unit. Pediatr Clin 60(3):581–592

38. McKinstry B, Hammersley V, Burton C, Pinnock H, Elton R, Dowell J, Sawdon N, Heaney D, Elwyn G, Sheikh A (2010) The quality, safety and content of telephone and face-to-face consultations: a comparative study. BMJ Quality & Safety 19(4):298–303

39. Mechanic OJ, Kimball AB (2019) Telehealth Systems.. StatPearls. Treasure Island (FL): StatPearls Publishing

40. Ozen A, Balasubramanian H (2013) The impact of case-mix on timely access to appointments in a primary care group practice. Health Care Management Science 16(2):101–118

41. Partnership to Fight Chronic Disease (2017). The Growing Crisis of Chronic Disease in the United States. Available at: https://www.fightchronicdisease.org/sites/default/files/docs/GrowingCrisisofChronicDiseaseintheUSfactsheet81009.pdf. Accessed June 2020

42. Perednia DA, Allen A (1995) Telemedicine technology and clinical applications. Jama 273(6):483–488

43. Pham H, Wang H (2000) Optimal (T, t) opportunistic maintenance of ak-out-of-n: G system with imperfect pm and partial failure. Naval Research Logistics (NRL) 47(3):223–239

44. Russo JE, McCool RR, Davies L (2016) VA Telemedicine: an analysis of cost and time savings. Telemedicine and e-Health 22(3):209–215

45. The Physicians Foundation (2018). 2018 survey of American physicians: practice patterns and perspectives. Available at: https://physiciansfoundation.org/research-insights/the-physicians-foundation-2018physician-survey. Accessed June 2020

46. The United States Census Bureau (2018). Population and housing unit estimates. Available at: https://www.census.gov/programs-surveys/popest.html?intcmp=serp. Accessed June 2020

47. Valero M, Arredondo M, Del Nogal F, Rodriguez J, Frias E (2000) Patient satisfaction with a home televisiting service based on interactive television over a cable network. J Telemed Telecare 6(1 suppl):99–101

48. Wu S-Y, Green A (2000) Projection of chronic illness prevalence and cost inflation. Santa Monica, CA:RAND Health, pp 18

49. Wu S-Y, Green A (2018) Global telemedicine market outlook 2022. Research and Markets

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.