This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Electronic Visits in Primary Care: Modeling, Analysis, and Scheduling Policies

Xiang Zhong, Jingshan Li, *Senior Member, IEEE*, Philip A. Bain, and Albert J. Musa

***Abstract*— Primary care, the backbone of the nation's healthcare system, is at the risk of collapse. Patients are dissatisfied due to poor access to care, and physicians are unhappy and burning out with an enormous amount of tasks. To improve the primary care access, many healthcare organizations have introduced electronic visits (or e-visits) to provide patient–physician communications through securing messages. In this paper, we introduce an analytical model to study e-visits in primary care clinics. Analytical formulas to evaluate the mean and variance of the patient length of visit in primary care clinics with e-visits are derived. System properties are investigated. In addition, comparisons of different scheduling policies between the office and the e-visits are carried out. The first come first serve, preemptive-resume, and non-preemptive policies are studied and the results show that the first come first serve policy typically leads to the best performance.**

***Note to Practitioners*—The primary care delivery system is under a lot of strain. Due to population growth and aging, and the expanded healthcare insurance coverage, the demand for primary care services has increased substantially in the past years. Patients have difficulty of getting timely access to care, while primary care physicians are facing insurmountable tasks. Electronic visit, or e-visit, as an alternative to the traditional office visit, provides an innovative way of patient–physician communication through securing messages. The successful implementation of e-visit relies on a proper understanding of the impact of e-visit on care access, and an appropriate design and scheduling of workforce and operations. Therefore, the objective of this paper is to develop an analytical model of the primary care delivery with e-visits, using which one can investigate the impact of e-visits on patient accessibility. In particular, the average value and variance of patients' length of visit for their encounters are evaluated. Different policies for physicians to schedule office and e-visit patients are compared. In addition, physicians' nondirect care activities, such as billings and documentations, are also considered in the model.**

***Index Terms*— E-visit, length of visit, monotonicity, patient flow, primary care, scheduling policy.**

I. Introduction

PRIMARY care, which is the backbone of the nation's healthcare system, is at a grave risk of collapse and facing a confluence of factors that could spell disaster [1], [2]. The patients are dissatisfied and have difficulty of getting timely access, while the physicians are unhappy with their jobs by facing insurmountable tasks. More patients need access to primary care but less medical students are choosing to enter the field. Recent studies have shown that 62 million people in the U.S. have no or inadequate access to primary care [3], but only 13% of the final-year medical students are planning on primary care careers [4]. The implementation of the Affordable Care Act will likely exacerbate the overcrowding in primary care clinics and the shortage of physicians [5]. Therefore, improving the accessibility of primary care is of significant importance.

The rapid development of information technology has made the delivery of healthcare over a distance possible, which introduces substantial opportunities. Many healthcare organizations have introduced online electronic visit programs, referred to as e-visit (or e-portal, e-service, and so on), to provide the patient–physician communication through securing messages [5]. Recent studies demonstrate that by introducing e-visits, significant savings can be obtained with improved access to care, and increased provider efficiency and patient satisfaction [6]–[10].

To better understand and implement e-visits, a mathematical model of primary care delivery through both the office and the e-visits is aspired. It can provide the care delivery process a fresh look from an integrated systems' engineering perspective. However, few quantitative models on e-visits are available in the current literature. How primary care physicians manage their operations in response to the introduction of e-visits is still an open question. Therefore, this paper is devoted to developing an analytical tool to investigate e-visit's impact on physician's practice, and identify the conditions that e-visits can improve patient accessibility.

As shown in Fig. 1, the care delivery process is essentially a service network and patients can get access to care through different venues: Web service, which is usually for patients to inquire some standard questions about simple diseases through an online questionnaire program; e-visits, mainly for the patients with low-acuity complaints and ongoing care of chronic diseases to communicate with physicians; office visits, traditional face-to-face encounters; urgent care, for after hour visits or walk-in for a quick treatment, where scheduling is not required; emergency department, for night and emergent visits. After finishing the online programs, a patient may still seek communication with his/her primary care physician through an e-visit if the online evaluation is not sufficient or satisfactory. In addition, the support staff will review the Web service results and, if needed, forward those complex inquiries to

Manuscript received December 16, 2015; revised March 28, 2016; accepted April 14, 2016. This paper was recommended for publication by Associate Editor C.-H. Chen and Editor M. P. Fanti upon evaluation of the reviewers' comments. This work was supported by the Division of Civil, Mechanical and Manufacturing Innovation under Grant CMMI-1233807 and Grant CMMI-1536987. *(Corresponding author: Jingshan Li.)*

X. Zhong and J. Li are with the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI 53706 USA (e-mail: xiang.zhong@wisc.edu; jingshan@engr.wisc.edu).

P. A. Bain and A. J. Musa are with Dean Health System, Inc., Madison, WI 53716 USA (e-mail: philip.bain@deancare.com; albert.musa@deancare.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2016.2555854

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

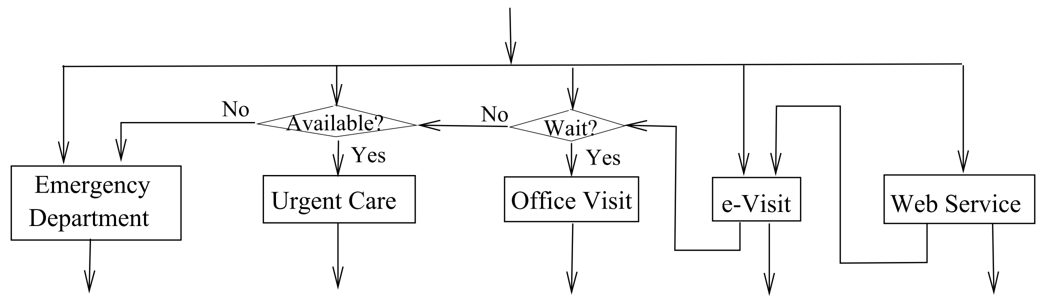

Fig. 1. Patient flow in primary care.

the patient's primary care physician for a follow-up e-visit. Therefore, patients can transfer from Web service to e-visit. Similarly, after e-visit, a patient might still need an office visit according to his/her health status and the complication of the disease. In the case of long queues or extended waiting time for office visits, or during after hours, patients may seek care services through other channels such as urgent care units, and if not available, emergency departments for prompt treatment.

Although electronic communication is desirable, the method to adopt it is still unsettled [11]. In particular, questions such as how is the workflow in primary care clinics affected by the use of e-visits and what is the impact of e-visits on resources to deliver proper care arise naturally. To answer these questions, the key is to evaluate the efficiency of primary care operations with e-visits and to determine the optimal scheduling policy coordinating office and e-visits. Unfortunately, the current literature lacks effective methods to address these issues. Computer simulation, as a prevailing tool to study healthcare delivery, such as primary care delivery, is often case study-based, and typically suffers from long model development and simulation times. To the best of our knowledge, only one analytical study exists, which analyzes primary care operations with e-visits and the focus is on identifying the incentives that drive the implementation of e-visits [12]. In addition, no effective method is available to address the unavailability of primary care providers due to other tasks on top of meeting with patients. Therefore, developing a novel method to model primary care delivery with e-visits, analyze its performance, and design the optimal operating policy is critically aspired, which is the goal of this paper.

The remainder of this paper is structured as follows. Section II reviews the related literature. Section III introduces the assumptions and formulates the problem. Performance analysis formulas are provided in Section IV. System monotonic properties for the mean and variance of the patient length of visit are discussed in Sections V and VI, respectively. Section VII devotes to the investigation of the scheduling policies coordinating office and e-visit services. Finally, conclusions are presented and the avenues for future research are highlighted in Section VIII. All the proofs are sketched in the Appendix.

II. Literature Review

Redesigning primary care clinics to improve the operational efficiency has been studied for decades. Most of the research addresses issues such as teamwork [13], [14], electronic health record and information systems [15], [16], medical homes [17], [18], payment systems [19], [20], and advanced access [21], [22]. For more details, see [23]–[25].

E-visit, as a novel alternative to the traditional office visit, has aroused growing attention in recent years. Many healthcare organizations, such as Henry Ford Health System, Mayo Clinic, Kaiser Permanente Health Plan, and the University of Pittsburgh Medical Center, have initiated e-visit programs [5], [6]–[10]. Most e-visit studies focus on investigating the effectiveness and patients/providers' experience of implementing e-visits. It is reported that the quality of care and the patient outcomes using e-visits are equivalent to those achieved with office visits [8], [9]. Implementing e-visits can free up extra office appointments for the patients with urgent and complicated issues, reduce urgent care and emergency room visits and inpatient hospital admissions, improve care for the senior population with chronic diseases, and substantially reduce the cost of care [6]–[10]. Additional studies investigate the issues such as billing and reimbursement, information system structures, legal and regulatory issues, financial return, and system implementation, and training [5]. As a quantitative analysis of e-visits, a patient health dynamics model is developed in [12] under the alternative primary care delivery mode, which includes the usage of e-visits and nonphysician providers. This paper quantifies the overall impact of adopting e-visits on physician's choices and expected earnings and patients' expected health outcomes. In a follow-up study based on these results, it is argued that e-visits provide a gateway for transforming traditional primary care delivery [26].

Discrete-event simulation has been used prevailingly to study primary care delivery (see [25], [27]–[31]), and questions pivoting around appointment scheduling, patient arrival, staffing allocation, and equipment maintenance in primary care clinic settings have been explored extensively. Analytical models, on the other hand, have less frequently been used to study primary care operations. Reviews of such models can be found in [30]–[33]. For instance, queuing models are introduced to determine the bed capacity and evaluate the patient cycle times in urgent care and maternity facilities in [34] and [35], respectively. Markov chain models are used to study the workflows in computed tomography test centers, gastroenterology clinics, and in-room care delivery systems [36]–[38]. A recursive procedure to address the limited availability of care providers with an application in a mammography imaging center is presented in [39]. The issues of outpatient appointment scheduling are studied in [40]–[42]. However, all these

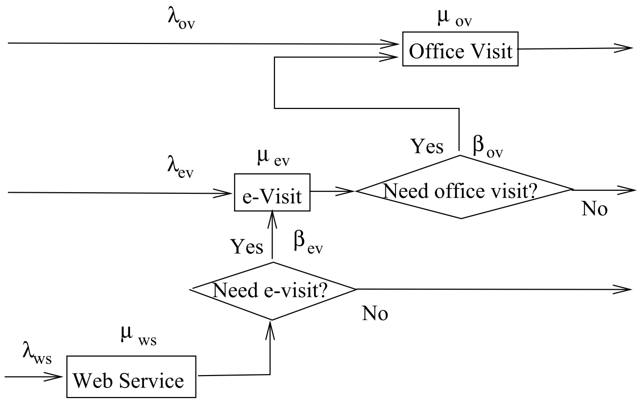

Fig. 2. Network model for primary care patient flow.

papers lack the specificity to address e-visit issues, and when evaluating the providers' productivity, the scenario that care providers may not be available for clinical service due to other duties is overlooked.

In spite of these efforts, no analytical study on patient flow and operations management has been carried out for primary care clinics with e-visits, and this paper intends to contribute to this end.

III. System Description and Modeling

As the focus of this paper is on studying e-visit and its impact on primary care physicians' operations, only Web services, e-visits, and office visits in a primary care delivery system are considered (see Fig. 2). The majority of patients in primary care clinics are associated with their dedicated primary care physician. Therefore, we consider a model with all the services linked with one physician. The following assumptions address the patients, the services, and their interactions.

1) The patients associated with the same primary care physician access care services with the following Poisson arrival rates: λ_{ws} for Web services, λ_{ev} for e-visits, and λ_{ov} for office visits.
2) The primary care physician's service times for e-visits and office visits are described by probability distributions with service rates μ_{ev} and μ_{ov}, coefficients of variation (CVs) cv_{ev} and cv_{ov}, as well as the third moments (or skewness) $E(S_{\text{ev}}^3)$ and $E(S_{\text{ov}}^3)$, correspondingly.
3) After Web service, a patient has the probability β_{ev} to seek an e-visit for further inquiries. After e-visit, a patient may need to go for an office visit with the probability β_{ov}.
4) The physician also deals with billings and documentations intermediately between patient visits. When no patient is waiting, he/she works on nondirect care related tasks, and the duration of tasks follows a probability distribution with the vacation rate μ_{υ}, the CV μ_{υ}, and the third moment $E(S_{\upsilon}^3)$. The physician will return to serve patients only after finishing an ongoing activity.
5) The following scheduling policies for coordinating office and e-visits are proposed: 1) non-preemptive, i.e., an ongoing e-visit service will not be interrupted even if an office visit patient arrives; 2) preemptive-resume, i.e., the current e-visit service can be interrupted if an office visit patient arrives, and the e-visit will resume afterward (in both the policies, office visit has a higher priority); and 3) first come first serve, i.e., the service will be carried out without priority but only based on who comes earlier.

In an appropriately defined state space, the system with assumptions 1)–5) forms a stationary random process. Note that a patient who finishes an e-visit still needs to go through the regular scheduling process for a subsequent office visit. Thus, from a physician's point of view, the combined arrival process can still be modeled as a stationary Poisson process. To quantify the system performance, an extensively used measure is the patient length of visit, which characterizes the duration of an episode of clinic stay [43]. However, a desired mean time performance alone cannot guarantee patient satisfaction—a large variation implies that some patients still wait for an extremely long time and even the mean waiting time is moderate. Moreover, unexpected variations may also impact the clinical outcome and patient safety [43]. Therefore, evaluating the variability of the patient length of visit is also important. Let T_i and Var_i denote the mean and variance of patient length of visit for the type i service, and $i = \text{ev}, \text{ov}$, representing the e-visit and the office visit. In the framework of 1)–5), T_i and Var_i are the functions of all system parameters

$$T_i = f_{T,i}(\mathcal{L}, \mathcal{M}, \mathcal{B}, \mathcal{CV}), \quad i = \text{ev}, \text{ov} \tag{1}$$
$$\text{Var}_i = f_{\text{Var},i}(\mathcal{L}, \mathcal{M}, \mathcal{B}, \mathcal{CV}, \mathcal{E}), \quad i = \text{ev}, \text{ov} \tag{2}$$

where

$$\begin{aligned}
\mathcal{L} &= [\lambda_{\text{ws}}, \lambda_{\text{ev}}, \lambda_{\text{ov}}] \\
\mathcal{M} &= [\mu_{\text{ev}}, \mu_{\text{ov}}, \mu_{\upsilon}] \\
\mathcal{B} &= [\beta_{\text{ev}}, \beta_{\text{ov}}] \\
\mathcal{CV} &= [\text{cv}_{\text{ev}}, \text{cv}_{\text{ov}}, \text{cv}_{\upsilon}] \\
\mathcal{E} &= \left[E\left(S_{\text{ev}}^3\right), E\left(S_{\text{ov}}^3\right), E\left(S_{\upsilon}^3\right)\right].
\end{aligned}$$

Remark 1: In addition to serving office and e-visit patients, physicians work on other tasks not directly encountering patients, such as documentation, paperwork, and dealing with insurance and billings. As summarized in [44], these nondirect care activities have become a significant part of physicians' workload. Assumption 4) implies that the physician works on these activities whenever no patients are waiting. When a new patient arrives, he/she will go to serve that patient after finishing the current activity.

The problem addressed in this paper is: under assumptions 1)–5), develop a method to evaluate the mean and variance of the patient length of visit, and investigate system properties and the impact of different scheduling policies between the office and the e-visits.

The solutions to this problem are presented in Sections IV–VII.

IV. Performance Evaluation

A. Average Length of Visit

Consider the primary care physician's operations described in Section III. For e-visit patients, the arrival includes the patients directly seeking e-visits and those coming to e-visits after Web services, which is characterized by the transition probability β_{ev}. Thus, the effective arrival rate for e-visits is

$$\lambda'_{\mathrm{ev}} = \lambda_{\mathrm{ev}} + \beta_{\mathrm{ev}}\lambda_{\mathrm{ws}}. \tag{3}$$

Similarly, the actual arrival rate for office visits also includes the patients who directly request office visits and those seeking face-to-face encounters after e-visits, which has the probability β_{ov}. Thus

$$\lambda'_{\mathrm{ov}} = \lambda_{\mathrm{ov}} + \beta_{\mathrm{ov}}\lambda'_{\mathrm{ev}}. \tag{4}$$

Remark 2: Note that, here, we assume the Web service patients who require additional care will first seek e-visits then, if needed, office visits. Clearly, there exists the possibility that they may directly go for office visits. Then, (4) can be revised as follows:

$$\lambda'_{\mathrm{ov}} = \lambda_{\mathrm{ov}} + \beta_{\mathrm{ov}}\lambda'_{\mathrm{ev}} + \beta'_{\mathrm{ov}}\lambda_{\mathrm{ws}}$$

where β'_{ov} is the proportion of Web service patients who need further services and directly go to office visits.

Since we focus on one physician and his/her patients, we can model the physician's activities as a single server working on two types of patients: office and e-visits. The nondirect care activities carried out by the physician when no patients are waiting can be modeled as vacations with a random vacation time. In addition, as the primary care physician offers both office and e-visit services, it is quite common that he/she may expect both the types of patients waiting to be served. Then how he/she schedules the work is of interest. Here, we consider three scheduling policies (assumption 5): non-preemptive, preemptive-resume, and first come first serve.

To derive the patient length of visit under non-preemptive and preemptive-resume policies, we first consider the case without vacations to obtain the patient waiting time as a function of the residual time serving each type of patients. Then, to incorporate vacations, the residual time is modified by accounting for vacation time when the server is idle. In this way, the adjusted waiting time is obtained and the patient length of visit can be calculated. In the case of the first come first serve policy, a new probability distribution can be constructed to model the physician's services. Similar approaches are applied to calculate the residual time and the waiting time.

To simplify the derivation and expressions, we introduce the following notations:

$$\rho_i = \frac{\lambda'_i}{\mu_i}, \quad i = \mathrm{ev}, \mathrm{ov} \tag{5}$$

$$\rho = \rho_{\mathrm{ev}} + \rho_{\mathrm{ov}} \tag{6}$$

$$\delta_i = \frac{1 + \mathrm{cv}_i^2}{2}, \quad i = \mathrm{ev}, \mathrm{ov}, \upsilon \tag{7}$$

$$\tau_i = \frac{1}{\mu_i}, \quad i = \mathrm{ev}, \mathrm{ov}, \upsilon \tag{8}$$

$$\omega_i = \tau_i\delta_i = \frac{E(S_i^2)}{2E(S_i)}, \quad i = \mathrm{ev}, \mathrm{ov}, \upsilon \tag{9}$$

where ρ and ρ_i's characterize the server utilization; and for the type i service, δ_i indicates the variability, τ_i represents the average time, and ω_i is a function of both the mean and the variability. In fact, ω_i represents the ratio between the second and first moments, multiplied by a factor of 0.5. In the case of exponential distributions, $\omega_i = \tau_i$, this variable can be explicitly explained by the mean service/vacation time.

Adopting these notations, as well as λ'_{ev} and λ'_{ov} introduced in (3) and (4), Theorem 1 provides the formulas to evaluate the average length of visit for office and e-visit patients with random service and vacation times.

Theorem 1: Under assumptions 1)–5), patient average lengths of visit for office and e-visit encounters can be calculated as follows:

$$T_{\mathrm{ev}} = \begin{cases} \dfrac{\rho_{\mathrm{ov}}\omega_{\mathrm{ov}} + \rho_{\mathrm{ev}}\omega_{\mathrm{ev}} + (1-\rho)\omega_\upsilon}{(1-\rho_{\mathrm{ov}})(1-\rho)} + \tau_{\mathrm{ev}}, & \text{non-preemptive policy} \\ \dfrac{\rho_{\mathrm{ov}}\omega_{\mathrm{ov}} + \rho_{\mathrm{ev}}\omega_{\mathrm{ev}} + (1-\rho)\omega_\upsilon}{(1-\rho_{\mathrm{ov}})(1-\rho)} + \dfrac{\tau_{\mathrm{ev}}}{1-\rho_{\mathrm{ov}}}, & \text{preemptive-resume policy} \\ \dfrac{\rho_{\mathrm{ov}}\omega_{\mathrm{ov}} + \rho_{\mathrm{ev}}\omega_{\mathrm{ev}}}{1-\rho} + \omega_\upsilon + \tau_{\mathrm{ev}}, & \text{first come first serve policy} \end{cases} \tag{10}$$

$$T_{\mathrm{ov}} = \begin{cases} \dfrac{\rho_{\mathrm{ov}}\omega_{\mathrm{ov}} + \rho_{\mathrm{ev}}\omega_{\mathrm{ev}} + (1-\rho)\omega_\upsilon}{1-\rho_{\mathrm{ov}}} + \tau_{\mathrm{ov}}, & \text{non-preemptive policy} \\ \dfrac{\rho_{\mathrm{ov}}\omega_{\mathrm{ov}} + (1-\rho)\omega_\upsilon}{1-\rho_{\mathrm{ov}}} + \tau_{\mathrm{ov}}, & \text{preemptive-resume policy} \\ \dfrac{\rho_{\mathrm{ov}}\omega_{\mathrm{ov}} + \rho_{\mathrm{ev}}\omega_{\mathrm{ev}}}{1-\rho} + \omega_\upsilon + \tau_{\mathrm{ov}}, & \text{first come first serve policy.} \end{cases} \tag{11}$$

Proof: See the Appendix. ■

When the service times and vacation time are exponentially distributed, all the values of cv_i's will be equal to 1. The expressions for office and e-visits' average lengths of visit can be simplified.

Corollary 1: Under assumptions 1)–5), with exponential service and vacation time distributions, patients' average lengths of visit for office and e-visit encounters can be calculated as

$$T_{\mathrm{ev}}^{\exp} = \begin{cases} \dfrac{\rho_{\mathrm{ov}}\tau_{\mathrm{ov}} + [1-\rho_{\mathrm{ov}}(2-\rho)]\tau_{\mathrm{ev}} + (1-\rho)\tau_\upsilon}{(1-\rho_{\mathrm{ov}})(1-\rho)}, & \text{non-preemptive policy} \\ \dfrac{\rho_{\mathrm{ov}}\tau_{\mathrm{ov}} + (1-\rho_{\mathrm{ov}})\tau_{\mathrm{ev}} + (1-\rho)\tau_\upsilon}{(1-\rho_{\mathrm{ov}})(1-\rho)}, & \text{preemptive-resume policy} \\ \dfrac{\rho_{\mathrm{ov}}\tau_{\mathrm{ov}} + (1-\rho_{\mathrm{ov}})\tau_{\mathrm{ev}}}{1-\rho} + \tau_\upsilon, & \text{first come first serve policy} \end{cases} \tag{12}$$

$$T_{\mathrm{ov}}^{\exp} = \begin{cases} \dfrac{\tau_{\mathrm{ov}} + \rho_{\mathrm{ev}}\tau_{\mathrm{ev}} + (1-\rho)\tau_\upsilon}{1-\rho_{\mathrm{ov}}}, & \text{non-preemptive policy} \\ \dfrac{\tau_{\mathrm{ov}} + (1-\rho)\tau_\upsilon}{1-\rho_{\mathrm{ov}}}, & \text{preemptive-resume policy} \\ \dfrac{(1-\rho_{\mathrm{ev}})\tau_{\mathrm{ov}} + \rho_{\mathrm{ev}}\tau_{\mathrm{ev}}}{1-\rho} + \tau_\upsilon, & \text{first come first serve policy.} \end{cases} \tag{13}$$

Proof: By plugging in $\omega_i = (1/\mu_i) = \tau_i$, $i = \text{ev}, \text{ov}, \upsilon$, (12) and (13) are obtained after several steps of algebraic operations. ■

B. Variance of Length of Visit

In this section, the variances of office and e-visit patients' lengths of visit under the three scheduling policies introduced in assumption 5) are evaluated.

Theorem 2: Under assumptions 1)–5), the variances of patient length of visit can be calculated as follows:

- *Non-preemptive Policy*

$$\begin{aligned}\text{Var}_{\text{ov}} &= \frac{2\delta_{\text{ov}}-1}{\mu_{\text{ov}}^2} + \frac{\rho_{\text{ov}}^2\omega_{\text{ov}}^2 - [(1-\rho)\omega_\upsilon + \rho_{\text{ev}}\omega_{\text{ev}}]^2}{(1-\rho_{\text{ov}})^2} \\ &\quad + \frac{(1-\rho)E(S_\upsilon^3)\mu_\upsilon + \lambda'_{\text{ov}}E(S_{\text{ov}}^3) + \lambda'_{\text{ev}}E(S_{\text{ev}}^3)}{3(1-\rho_{\text{ov}})}\end{aligned} \tag{14}$$

$$\begin{aligned}\text{Var}_{\text{ev}} &= \frac{2\delta_{\text{ev}}-1}{\mu_{\text{ev}}^2} + \frac{(1-\rho)\omega_\upsilon + \rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}}}{(1-\rho_{\text{ov}})^2(1-\rho)} \\ &\quad \cdot \left(\frac{2\rho_{\text{ov}}\omega_{\text{ov}}}{1-\rho_{\text{ov}}} + \frac{\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}}}{1-\rho} - \omega_\upsilon\right) \\ &\quad + \frac{(1-\rho)E(S_\upsilon^3)\mu_\upsilon + \lambda'_{\text{ov}}E(S_{\text{ov}}^3) + \lambda'_{\text{ev}}E(S_{\text{ev}}^3)}{3(1-\rho_{\text{ov}})^2(1-\rho)}.\end{aligned} \tag{15}$$

- *Preemptive-Resume Policy*

$$\begin{aligned}\text{Var}_{\text{ov}} &= \frac{2\delta_{\text{ov}}-1}{\mu_{\text{ov}}^2} + \frac{\rho_{\text{ov}}^2\omega_{\text{ov}}^2 - (1-\rho)^2\omega_\upsilon^2}{(1-\rho_{\text{ov}})^2} \\ &\quad + \frac{(1-\rho)E(S_\upsilon^3)\mu_\upsilon + \lambda'_{\text{ov}}E(S_{\text{ov}}^3)}{3(1-\rho_{\text{ov}})}\end{aligned} \tag{16}$$

$$\begin{aligned}\text{Var}_{\text{ev}} &= \frac{2\delta_{\text{ev}}-1}{\mu_{\text{ev}}^2(1-\rho_{\text{ov}})^2} + \frac{(1-\rho)\omega_\upsilon + \rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}}}{(1-\rho_{\text{ov}})^2(1-\rho)} \\ &\quad \cdot \left(\frac{2\rho_{\text{ov}}\omega_{\text{ov}}}{1-\rho_{\text{ov}}} + \frac{\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}}}{1-\rho} - \omega_\upsilon\right) \\ &\quad + \frac{(1-\rho)E(S_\upsilon^3)\mu_\upsilon + \lambda'_{\text{ov}}E(S_{\text{ov}}^3) + \lambda'_{\text{ev}}E(S_{\text{ev}}^3)}{3(1-\rho_{\text{ov}})^2(1-\rho)}.\end{aligned} \tag{17}$$

- First Come First Serve Policy

$$\begin{aligned}\text{Var}_j &= \frac{2\delta_j - 1}{\mu_j^2} + \frac{(\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}})^2}{(1-\rho)^2} - \omega_\upsilon^2 \\ &\quad + \frac{\mu_\upsilon}{3}E(S_\upsilon^3) \\ &\quad + \frac{\lambda'_{\text{ov}}E(S_{\text{ov}}^3) + \lambda'_{\text{ev}}E(S_{\text{ev}}^3)}{3(1-\rho)}, \quad j = \text{ev}, \text{ov}.\end{aligned} \tag{18}$$

Proof: See the Appendix. ■

As one can see, in addition to the mean and CV of service and vacation times, the third moments play a role in determining the variation of the system cycle time. When exponential service and vacation time distributions are assumed, all the CVs are equal to 1 and the expressions for the third moments can be explicitly expressed by parameter μ_j's.

Corollary 2: Under assumptions 1)–5), the variances of patient length of visit under exponential service and vacation time distributions can be calculated as follows:

- *Non-preemptive Policy*

$$\begin{aligned}\text{Var}_{\text{ov}}^{\text{exp}} &= \frac{1}{\mu_{\text{ov}}^2} - \frac{\rho_{\text{ev}}^2}{(1-\rho_{\text{ov}})^2\mu_{\text{ev}}^2} + \frac{\rho_{\text{ov}}^2}{(1-\rho_{\text{ov}})^2\mu_{\text{ov}}^2} \\ &\quad + \frac{(1-\rho)(1+\rho_{\text{ev}}-\rho_{\text{ov}})}{(1-\rho_{\text{ov}})^2\mu_\upsilon^2} + \frac{2\rho_{\text{ov}}}{(1-\rho_{\text{ov}})\mu_{\text{ov}}^2} \\ &\quad + \frac{2\rho_{\text{ev}}}{(1-\rho_{\text{ov}})\mu_{\text{ev}}^2} - \frac{2\rho_{\text{ev}}(1-\rho)}{(1-\rho_{\text{ov}})^2\mu_{\text{ev}}\mu_\upsilon}\end{aligned} \tag{19}$$

$$\begin{aligned}\text{Var}_{\text{ev}}^{\text{exp}} &= \frac{1}{\mu_\upsilon^2(1-\rho_{\text{ov}})^2} + \frac{\rho_{\text{ev}}^2}{\mu_{\text{ev}}^2(1-\rho)^2(1-\rho_{\text{ov}})^2} \\ &\quad + \frac{2\rho_{\text{ev}}}{\mu_{\text{ev}}^2(1-\rho)(1-\rho_{\text{ov}})^2} + \frac{\rho_{\text{ov}}^2(3-2\rho-\rho_{\text{ov}})}{\mu_{\text{ov}}^2(1-\rho)^2(1-\rho_{\text{ov}})^3} \\ &\quad + \frac{2\rho_{\text{ov}}}{\mu_\upsilon\mu_{\text{ov}}(1-\rho_{\text{ov}})^3} + \frac{2\rho_{\text{ov}}}{\mu_{\text{ov}}^2(1-\rho)(1-\rho_{\text{ov}})^2} \\ &\quad + \frac{2\rho_{\text{ov}}\rho_{\text{ev}}(2-\rho_{\text{ov}}-\rho)}{\mu_{\text{ov}}\mu_{\text{ev}}(1-\rho)^2(1-\rho_{\text{ov}})^3} + \frac{1}{\mu_{\text{ev}}^2}.\end{aligned} \tag{20}$$

- *Preemptive-Resume Policy*

$$\begin{aligned}\text{Var}_{\text{ov}}^{\text{exp}} &= \frac{\rho_{\text{ov}}^2}{\mu_{\text{ov}}^2(1-\rho_{\text{ov}})^2} + \frac{(1-\rho)(1+\rho_{\text{ev}}-\rho_{\text{ov}})}{\mu_\upsilon^2(1-\rho_{\text{ov}})^2} \\ &\quad + \frac{2\rho_{\text{ov}}}{\mu_{\text{ov}}^2(1-\rho_{\text{ov}})} + \frac{1}{\mu_{\text{ov}}^2}\end{aligned} \tag{21}$$

$$\begin{aligned}\text{Var}_{\text{ev}}^{\text{exp}} &= \frac{1}{\mu_\upsilon^2(1-\rho_{\text{ov}})^2} + \frac{\rho_{\text{ev}}^2}{\mu_{\text{ev}}^2(1-\rho)^2(1-\rho_{\text{ov}})^2} \\ &\quad + \frac{2\rho_{\text{ev}}}{\mu_{\text{ev}}^2(1-\rho)(1-\rho_{\text{ov}})^2} + \frac{\rho_{\text{ov}}^2(3-2\rho-\rho_{\text{ov}})}{\mu_{\text{ov}}^2(1-\rho)^2(1-\rho_{\text{ov}})^3} \\ &\quad + \frac{2\rho_{\text{ov}}}{\mu_\upsilon\mu_{\text{ov}}(1-\rho_{\text{ov}})^3} + \frac{2\rho_{\text{ov}}}{\mu_{\text{ov}}^2(1-\rho)(1-\rho_{\text{ov}})^2} \\ &\quad + \frac{2\rho_{\text{ov}}\rho_{\text{ev}}(2-\rho_{\text{ov}}-\rho)}{\mu_{\text{ov}}\mu_{\text{ev}}(1-\rho)^2(1-\rho_{\text{ov}})^3} + \frac{1}{(1-\rho_{\text{ov}})^2\mu_{\text{ev}}^2}.\end{aligned} \tag{22}$$

- *First Come First Serve Policy*

$$\text{Var}_j^{\text{exp}} = \frac{1}{\mu_j^2} + \frac{1}{\mu_\upsilon^2} + \frac{2\left(\frac{\rho_{\text{ov}}}{\mu_{\text{ov}}^2} + \frac{\rho_{\text{ev}}}{\mu_{\text{ev}}^2}\right)}{(1-\rho)} + \frac{\left(\frac{\rho_{\text{ov}}}{\mu_{\text{ov}}} + \frac{\rho_{\text{ev}}}{\mu_{\text{ev}}}\right)^2}{(1-\rho)^2}, \quad j = \text{ov}, \text{ev}. \tag{23}$$

Proof: By plugging in $\delta_i = 1$, $\omega_i = (1/\mu_i)$, and $E(S_i^3) = (6/\mu_i^3)$, $i = \text{ev}, \text{ov}, \upsilon$, (19)–(23) can be obtained after several steps of algebraic operations. ■

Building upon these system performance evaluation formulas, system properties like monotonicity can be studied. Then, questions such as how do system parameters impact performance measures and what are the directions to improve system performance can be answered. In Sections V and VI, the properties of the mean and variance of lengths of visit are discussed, and different scheduling policies are compared.

V. Property of Average Length of Visit

In this section, we investigate the impact of routing probabilities on e-visit and office visit patients' average lengths of visit. Since $\beta_{\rm ev}$ and $\beta_{\rm ov}$ are the probabilities that patients continue to seek e-visits and office visits after Web services and e-visits, the monotonicity of T_i, $i = {\rm ev}, {\rm ov}$ with respect to $\beta_{\rm ev}$ and $\beta_{\rm ov}$ could provide insights on how e-visits impact patient access to primary care.

A. Monotonicity of T_{ov} With Respect to β_{ov}

Proposition 1: Under assumptions 1)–5), $T_{\rm ov}$ is monotonically increasing with respect to $\beta_{\rm ov}$, i.e., $(\partial T_{\rm ov}/\partial \beta_{\rm ov}) > 0$, if and only if

$$\begin{cases} \omega_{\rm ov} > (\omega_v - \omega_{\rm ev})\rho_{\rm ev}, & \text{non-preemptive policy} \\ \omega_{\rm ov} > \omega_v \rho_{\rm ev}, & \text{preemptive-resume policy} \\ \text{without condition,} & \text{first come first serve policy.} \end{cases}$$

Proof: See the Appendix. ■

Intuitively, if the routing probability of seeking office visits after e-visits, $\beta_{\rm ov}$, is increasing, the physician's workload with office visit patients is increasing. Under the non-preemptive policy, when $\omega_{\rm ov} > (\omega_v - \omega_{\rm ev})\rho_{\rm ev}$, the office visit patient's length of visit will increase with respect to $\beta_{\rm ov}$ and will be nonincreasing vice versa. Such a condition suggests that, roughly, the moment ratio of the office service is larger than that of the difference between vacation and e-visit.

In practice, this type of condition typically holds, since both the e-visit and the vacation have lower priorities than the office visit and usually take a shorter time compared with the office visit. The difference will be even smaller considering the discount factor $\rho_{\rm ev} < 1$. In particular, when service and vacation times are exponentially distributed, this condition is simplified to $\tau_{\rm ov} > (\tau_v - \tau_{\rm ev})\rho_{\rm ev}$, which again holds most of the time.

Under the preemptive-resume policy, the condition becomes more strict, where $\omega_{\rm ov} > \omega_v \rho_{\rm ev}$ (in the exponential case, $\tau_{\rm ov} > \tau_v \rho_{\rm ev}$) is required. The reason is that under the preemptive-resume policy, the physician will stop working on e-visit patients and immediately serve an incoming office visit patient. Then, the service time and the variability of e-visits will not play a significant role in the waiting time of office visit patients compared with the non-preemptive case, where the physician has to finish any ongoing e-visit service before moving to office visit patients. However, since vacations usually take a shorter time and $\rho_{\rm ev} < 1$, this condition is typically satisfied, so that the monotone increasing property holds.

An illustration of such monotonicity property in exponential scenarios is shown in Fig. 3, in which the parameters are selected as follows:

$$\beta_{\rm ov} \in [0, 0.5), \qquad \beta_{\rm ev} = 0.5 \tag{24}$$
$$\text{Case A:} \qquad \tau_v = 30\tau_{\rm ev} = 10\tau_{\rm ov} \tag{25}$$
$$\text{Case B:} \qquad \tau_v = \tau_{\rm ev} = \frac{\tau_{\rm ov}}{3}. \tag{26}$$

The reason to include the seldom occurring Case A is to show the decreasing monotonicity. As one can see, when office visits take a longer time, which meets the requirement

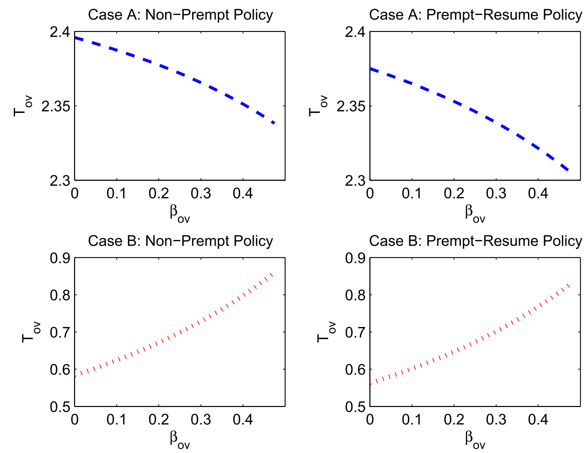

Fig. 3. Monotonicity of $T_{\rm ov}$ with respect to $\beta_{\rm ov}$.

in Proposition 1, $T_{\rm ov}$ is increasing with respect to $\beta_{\rm ov}$ (Case B). However, if vacation (or nondirect care) takes an extremely long time than office and e-visits, $T_{\rm ov}$ could decrease with respect to $\beta_{\rm ov}$ (Case A). In a sense, waiting for short office visits is better than for long vacations.

When the first come first serve policy is applied, the office visit patient's length of visit is monotonically increasing with respect to $\beta_{\rm ov}$ without any condition. In this case, both the office and the e-visits are treated with equal priority. Increasing physician's workload [$\rho_{\rm ov}$ and ρ in (11) and (13)] will lead to a longer patient length of visit.

Therefore, in most of the practical cases, if more patients need to seek additional office visits after e-visits, the accessibility to office visits can be further impaired. Thus, the method to implement e-visits to limit this routing probability is of importance, and will be part of future work.

B. Monotonicity of T_{ov} With Respect to β_{ev}

Proposition 2: Under assumptions 1)–5), $T_{\rm ov}$ is monotonically increasing with respect to $\beta_{\rm ev}$, i.e., $(\partial T_{\rm ov}/\partial \beta_{\rm ev}) > 0$, if and only if

$$\begin{cases} \beta_{\rm ov}\mu_{\rm ev}\omega_{\rm ov} > (\mu_{\rm ov} - \lambda_{\rm ov})(\omega_v - \omega_{\rm ev}), \\ \qquad\qquad\qquad\qquad \text{non-preemptive policy} \\ \beta_{\rm ov}\mu_{\rm ev}\omega_{\rm ov} > (\mu_{\rm ov} - \lambda_{\rm ov})\omega_v, \\ \qquad\qquad\qquad\qquad \text{preemptive-resume policy} \\ \text{without condition,} \qquad \text{first come first serve policy.} \end{cases}$$

Proof: See the Appendix. ■

Again, the increasing monotonicity exists without any condition under the first come first serve policy. For non-preemptive and preemptive-resume policies, the necessary and sufficient conditions become more complex.

When $\beta_{\rm ev}$ is increasing, i.e., more patients continue to seek e-visits after Web services, which leads to an increase in the number of patients to further come to the office visit (as $\beta_{\rm ov} > 0$ and is kept constant). Since $\beta_{\rm ev}$ mainly affects the arrival of e-visits, only when $\beta_{\rm ov}$ is large enough, the increase of follow-up office visits can exert a significant effect (which explains the conditions with the factor $\beta_{\rm ov}$ on the left-hand side of the inequalities in Proposition 2, required for both the policies).

For the non-preemptive policy, if the physician spends more time, which also has a higher variability on office and

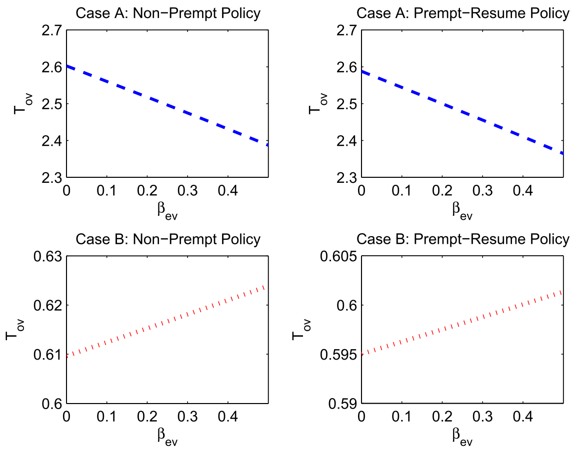

Fig. 4. Monotonicity of T_{ov} with respect to β_{ev}.

e-visits than vacations, a longer length of visit can be observed (which explains the condition regarding the ω_{ov} and $\omega_v - \omega_{\mathrm{ev}}$ factors in Proposition 2 for the non-preemptive policy). For the preemptive-resume policy, additional e-visit patients will not significantly impact office visits, since the physician will stop working on any e-visit and immediately work on the coming office visit patient. Thus, the condition in Proposition 2 for the preemptive-resume policy becomes stricter, where the $\omega_v - \omega_{\mathrm{ev}}$ term changes to ω_v.

Note that these conditions are necessary and sufficient, which indicates that if these conditions are not met, T_{ov} will be monotone nonincreasing with respect to β_{ev}. Fig. 4 shows such properties in exponential cases. System parameters are selected as in (25) and (26), but (24) is replaced by (27) to represent the scenario that the Web service has a higher referral ratio than e-visits

$$\beta_{\mathrm{ev}} \in [0, 0.95), \quad \beta_{\mathrm{ov}} = 0.1. \tag{27}$$

As exhibited in Fig. 4, when the vacation time is much longer, waiting for more office visits could be even beneficial, so that the decreasing monotonicity can be observed.

C. Monotonicity of T_{ev}

Unlike T_{ov}, the monotonicity of T_{ev} is consistent for the non-preemptive, preemptive-resume, and first come first serve policies.

Proposition 3: Under assumptions 1)–5), T_{ev} is monotonically increasing with respect to β_{ev} and β_{ov}, i.e., $(\partial T_{\mathrm{ev}}/\partial \beta_i) > 0$, $i = \mathrm{ov}, \mathrm{ev}$.

Proof: See the Appendix. ■

Proposition 3 articulates that the length of visit of e-visit patients is always monotonically increasing with respect to β_{ev} and β_{ov}, no matter which policy is implemented. Larger β_{ov} and β_{ev} increase the effective arrivals, resulting in more patients waiting in line. In close, under all the policies, a newly arrived e-visit patient needs to wait until all the types of patients in line are finished. Thus, the increase of average length of visit can be foreseen.

D. Discussion

The scheduling policies introduced in this paper can be categorized into two groups, based on whether the arriving patients are with or without priority. For the first come first policy, patient types are not differentiated, and increasing either β_{ov} or β_{ev} increases the total patient arrival, so does the server intensity ρ_{ov}, ρ_{ev}, and ρ. In addition, the effect of vacation on patient length of visit is independent of server intensity, which is elucidated in (10) and (11) (where the terms related to ω_v or τ_v are independent of ρ_{ev}, ρ_{ov}, and ρ). Therefore, it is straightforward that the increasing monotonicity holds for the lengths of visit of both the office and e-visit patients unconditionally.

For the policies with priorities, the results differ for the office and e-visit patients. As the e-visit patients have a lower priority, their waiting incorporates the waiting for all the patients in line and the waiting for the physician to return from a vacation. Larger β_{ev} or β_{ov} increases the overall patient arrival, and thus the overall number of patients waiting in line. Therefore, the monotonicity of their length of visit holds naturally without conditions.

On the other hand, office visit patients are mainly waiting for other office visit patients in line and the physician returning from a vacation. There exists a tradeoff between waiting for more office and e-visits due to the increase of β_{ov} or β_{ev} and waiting for potentially fewer vacations. Therefore, conditions are required to ensure the monotone increasing of the length of visit for office visits. In extreme cases, if vacations are very long or suffer large variations ($\omega_v \gg \omega_{\mathrm{ev}}$ or $\omega_v \gg \omega_{\mathrm{ov}}$), then having more office and e-visit arrivals could be beneficial (i.e., T_{ov} is monotonically decreasing with respect to β_{ov} and β_{ev}). Moreover, for T_{ov} to be monotonically increasing with β_{ev}, as β_{ev} mainly affects e-visits and its impact on office visit is through β_{ov}, additional conditions on β_{ov} are required.

The conditions for the preemptive-resume policy are always stricter than that for the non-preemptive policy. In the former case, physicians will stop the ongoing e-visit, and thus, only significant changes in e-visits will impose effects on office visits, while in the latter case, physicians will finish the current e-visit service, and any change in e-visits may immediately impact office visits.

In summary, in practical cases, office visits have a higher demand and take a longer time, and then both T_{ov} and T_{ev} are monotonically increasing with respect to β_{ov} and β_{ev}.

VI. Property of Variance of Length of Visit

A. Monotonicity of Var_{ov}

First, we investigate the monotonicity of variance of length of visit $\mathrm{Var}_{\mathrm{ov}}$ with respect to β_{ov}. The increasing monotonicity holds under a sufficient but not necessary condition.

Proposition 4: Under assumptions 1)–5), $\mathrm{Var}_{\mathrm{ov}}$ is monotonically increasing with respect to β_{ov}, i.e., $(\partial \mathrm{Var}_{\mathrm{ov}}/\partial \beta_{\mathrm{ov}}) > 0$, if

$$\begin{cases} \omega_{\mathrm{ov}} \geq \rho_{\mathrm{ev}}|\omega_v - \omega_{\mathrm{ev}}| \text{ and} \\ \mu_{\mathrm{ov}} E\left(S_{\mathrm{ov}}^3\right) + \lambda'_{\mathrm{ev}} E\left(S_{\mathrm{ev}}^3\right) \geq \rho_{\mathrm{ev}} \mu_v E\left(S_v^3\right), \\ \qquad \text{non-preemptive policy} \\ \omega_{\mathrm{ov}} \geq \rho_{\mathrm{ev}} \omega_v \text{ and } \mu_{\mathrm{ov}} E\left(S_{\mathrm{ov}}^3\right) \geq \rho_{\mathrm{ev}} \mu_v E\left(S_v^3\right), \\ \qquad \text{preemptive-resume policy} \\ \text{without condition, first come first serve policy.} \end{cases}$$

Proof: See the Appendix. ■

The sufficient conditions for the variance of length of visit are much more complex compared with that of the average length of visit, since the third moments are involved. These conditions indicate that when the office visit has a longer service time and a larger variance, and the vacation (i.e., nondirect care activity) has a smaller moment ratio, then more patients seeking office visits after e-visits will lead to a larger variability in the patient flow. Similar to the T_{ov} case, the sufficient conditions under the preemptive-resume policy are stricter than those under the non-preemptive policy. Under the first come first serve policy, fortunately, the monotonicity is straightforward that the variance of length of visit for office visit patients is always increasing when more patients shift to office visits.

It can be noticed that the characteristic of vacation plays an important role in determining the monotonicity of the system performance indices—as long as all the three distribution moments of vacation are small enough, the monotonicity holds. One other observation is that the length of visit variation is affected by multiple factors comprising the first, second, and third distribution moments. Therefore, the effect of each factor on determining the holistic system performance weakens. Inference can be drawn that the increase of the referral ratio escalates the variation of the system, but not in a strong manner as it impacts the mean time performance. Similar properties are witnessed for the variance with respect to β_{ev}.

These sufficient conditions are less complicated under exponential assumptions.

Corollary 3: Under assumptions 1)–5), with exponential service and vacation time distributions, $(\partial \text{Var}_{\text{ov}}^{\text{exp}}/\partial \beta_{\text{ov}}) > 0$, if

$$\begin{cases} \tau_{\text{ov}} \geq \rho_{\text{ev}}|\tau_{\upsilon} - \tau_{\text{ev}}| \text{ and} \\ \tau_{\text{ov}}^2 \geq \rho_{\text{ev}}\left(\tau_{\upsilon}^2 - \tau_{\text{ev}}^2\right) \\ \qquad\qquad\qquad \text{non-preemptive policy} \\ \tau_{\text{ov}}^2 \geq \rho_{\text{ev}}\tau_{\upsilon}^2, \\ \qquad\qquad\qquad \text{preemptive-resume policy.} \end{cases}$$

Proof: By plugging in $\delta_i = 1$, $\omega_i = (1/\mu_i)$, and $E(S_i^3) = (6/\mu_i^3)$, $i = \text{ev}, \text{ov}, \upsilon$, and eliminating the redundant conditions, the expressions above are obtained. ■

Conditions described above are typically met, since the office visit usually takes a longer time than e-visits and vacations, and $\rho_{\text{ev}} < 1$. Illustrations are shown in Fig. 5, where the same parameter settings (24)–(26) matching the conditions in Corollary 3 are used. When office visits take a longer time, Var_{ov} increasing with β_{ov} is observed (Case B). However, when the vacation is abnormally long, the decreasing phenomenon can be observed (Case A).

Next we study the monotonicity of Var_{ov} with respect to β_{ev}. Similar to T_{ov}'s case, the sufficient conditions become stricter under non-preemptive and preemptive-resume policies, but no condition is demanded under the first come first policy.

Proposition 5: Under assumptions 1)–5), Var_{ov} is monotonically increasing with respect to β_{ev}, i.e., $(\partial \text{Var}_{\text{ov}}/\partial \beta_{\text{ev}}) > 0$ if

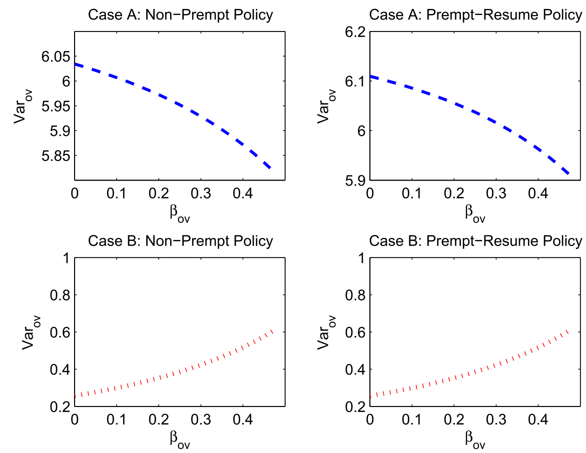

Fig. 5. Monotonicity of Var_{ov} with respect to β_{ov}.

$$\begin{cases} \beta_{\text{ov}}\mu_{\text{ev}}\omega_{\text{ov}} > (\mu_{\text{ov}} - \lambda_{\text{ov}})|\omega_{\text{ev}} - \omega_{\upsilon}| \\ \text{and } \mu_{\text{ev}}E\left(S_{\text{ev}}^3\right) \geq \mu_{\upsilon}E\left(S_{\upsilon}^3\right), \\ \qquad\qquad \text{non-preemptive policy} \\ \beta_{\text{ov}}\mu_{\text{ev}}\omega_{\text{ov}} > (\mu_{\text{ov}} - \lambda_{\text{ov}})\omega_{\upsilon} \\ \text{ and } \beta_{\text{ov}}\mu_{\text{ov}}\mu_{\text{ev}}E\left(S_{\text{ov}}^3\right) \geq (\mu_{\text{ov}} - \lambda_{\text{ov}})\mu_{\upsilon}E\left(S_{\upsilon}^3\right), \\ \qquad\qquad \text{preemptive-resume policy} \\ \text{without condition,} \quad \text{first come first serve policy.} \end{cases}$$

Proof: See the Appendix. ■

The sufficient conditions in Proposition 5 contain the necessary and sufficient conditions presented in Proposition 2. In addition to that, conditions regarding the third moments are required. Likewise, the conditions are more rigid for the preemptive-resume policy compared with the non-preemptive policy. The increasing monotonicity always holds for the first come first serve policy.

In exponential cases, conditions are simplified. Fig. 6 shows the trend of variance changing with the referral ratio β_{ev} under exponential settings.

Corollary 4: Under assumptions 1)–5), with exponential service and vacation time distributions, $(\partial \text{Var}_{\text{ov}}^{\text{exp}}/\partial \beta_{\text{ev}}) > 0$ if

$$\begin{cases} \beta_{\text{ov}}\mu_{\text{ev}}\tau_{\text{ov}} > (\mu_{\text{ov}} - \lambda_{\text{ov}})|\tau_{\text{ev}} - \tau_{\upsilon}| \\ \text{ and } \tau_{\text{ev}} \geq \tau_{\upsilon} \\ \qquad\qquad \text{non-preemptive policy} \\ \beta_{\text{ov}}\mu_{\text{ev}}\tau_{\text{ov}} > (\mu_{\text{ov}} - \lambda_{\text{ov}})\tau_{\upsilon} \\ \text{ and } \beta_{\text{ov}}\mu_{\text{ev}}\tau_{\text{ov}}^2 \geq (\mu_{\text{ov}} - \lambda_{\text{ov}})\tau_{\upsilon}^2, \\ \qquad\qquad \text{preemptive-resume policy.} \end{cases}$$

Proof: By plugging in $\delta_i = 1$, $\omega_i = (1/\mu_i)$, and $E(S_i^3) = (6/\mu_i^3)$, $i = \text{ev}, \text{ov}, \upsilon$, the expressions above can be obtained. ■

B. Monotonicity of Var_{ev}

First, we consider non-preemptive and preemptive-resume policies. The monotonicity conditions for Var_{ev} to increase with respect to β_{ov} and β_{ev} become much more complicated. Thus, only the sufficient conditions are pursued. Under both the policies, the same sufficient condition for Var_{ev} to be monotonically increasing with respect to β_{ov} and β_{ev} has been derived.

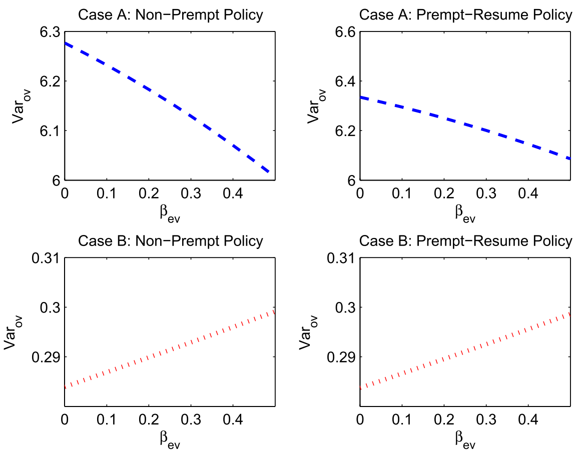

Fig. 6. Monotonicity of $\mathrm{Var}_{\mathrm{ov}}$ with respect to β_{ev}.

Proposition 6: Under assumptions 1)–5), with non-preemptive and preemptive-resume policies, if $\omega_{\mathrm{ov}} \geq (\omega_v/2)$, then $(\partial \mathrm{Var}_{\mathrm{ev}}/\partial \beta_i) > 0$, $i = \mathrm{ev}, \mathrm{ov}$, i.e., $\mathrm{Var}_{\mathrm{ev}}$ is monotonically increasing with respect to β_{ov} and β_{ev}.

Proof: See the Appendix. ■

Since office visits normally take a longer time and have a larger variability compared with vacations, $\omega_{\mathrm{ov}} \geq \frac{\omega_v}{2}$ usually holds and the sufficient condition is satisfied. For the first come first serve policy, monotonic properties with respect to both β_{ov} and β_{ev} hold unconditionally.

Proposition 7: Under assumptions 1)–5), with the first come first serve policy, $\mathrm{Var}_{\mathrm{ev}}$ is monotonically increasing with respect to β_{ov} and β_{ev}, that is

$$\frac{\partial \mathrm{Var}_{\mathrm{ev}}}{\partial \beta_{\mathrm{ov}}} > 0, \quad \frac{\partial \mathrm{Var}_{\mathrm{ev}}}{\partial \beta_{\mathrm{ev}}} > 0.$$

Proof: See the Appendix. ■

C. Discussion

For the first come first serve policy, similar to the property of the average length of visit, the terms (the first, second, and third moments) related to vacations are independent of the patient arrival [see (18)]. Thus, vacations would not affect the system monotonicity property regarding arrival, which is the only term that links to routing probabilities. For the non-preemptive and preemptive-resume policies, the sign of the term ω_v is negative, and the monotonicity of the coefficient of ω_v with respect to β_{ov} or β_{ev} is not clear [see (14)–(17)]. Therefore, the property of vacation plays an important role in determining the monotonicity of variation for office and e-visits.

In summary, the monotone increasing property of variance requires a short vacation time, less vacation variations, and a small third distribution moment of vacation. In most practical cases, the variances $\mathrm{Var}_{\mathrm{ov}}$ and $\mathrm{Var}_{\mathrm{ev}}$ are monotonically increasing with respect to β_{ov} and β_{ev}, as long as office visits need a longer time and have high variability.

VII. Comparison of Scheduling Policies

A. Average Length of Visit

To improve physicians' operations and design an efficient daily workflow, this section is dedicated to identifying the optimal scheduling policy and its conditions. By considering all the patients who need physicians' services, the overall patient average length of visit can be obtained, which is a weighted average length of visit of both the office and e-visit patients

$$T = p_{\mathrm{ev}} T_{\mathrm{ev}} + p_{\mathrm{ov}} T_{\mathrm{ov}} \tag{28}$$

$$p_i = \frac{\lambda_i'}{\lambda_{\mathrm{ev}}' + \lambda_{\mathrm{ov}}'}, \quad i = \mathrm{ev}, \mathrm{ov}. \tag{29}$$

Comparing the three scheduling policies, we have Proposition 8.

Proposition 8: Under assumptions 1)–5), assume $\mu_{\mathrm{ov}} < \mu_{\mathrm{ev}}$, then

$$T_{\text{Non-Preemp}} > T_{\text{FCFS}}$$

and in addition, if $\mathrm{cv}_{\mathrm{ev}} < 1$

$$T_{\text{Preemp}} > T_{\text{Non-Preemp}}$$

where T_{Preemp}, $T_{\text{Non-Preemp}}$, and T_{FCFS} are the weighted average lengths of visit under the preemptive-resume, non-preemptive, and first come first serve policies, respectively.

Proof: See the Appendix. ■

The second inequality in Proposition 8 indicates that it is always preferable to finish the current e-visit without interrupting the ongoing work, and then start working on office visit patients. The rationale behind this is that the preemptive-resume policy implies that the physician needs to restart the interrupted e-visit after finishing the office visits, which creates more variations for e-visits that can lead to a longer length of visit. In practice, office visits usually take a longer service time than e-visits ($\mu_{\mathrm{ov}} < \mu_{\mathrm{ev}}$), and along with the second inequality, it is justified that the first come first serve policy leads to the highest system productivity, i.e., the shortest average length of visit.

B. Variance of Length of Visit

The above analysis manifests that the non-preemptive policy is superior to the preemptive-resume policy in terms of mean time performance. To compare the variances of length of visit under different scheduling policies, define the weighted variance of length of visit of both the office and e-visit patients as

$$\mathrm{Var} = p_{\mathrm{ev}} \mathrm{Var}_{\mathrm{ev}} + p_{\mathrm{ov}} \mathrm{Var}_{\mathrm{ov}}. \tag{30}$$

Then, the following necessary and sufficient condition is obtained.

Proposition 9: Under assumptions 1)–5), if and only if

$$E(S_{\mathrm{ev}}^3) < \frac{1}{(1-\rho_{\mathrm{ov}})\mu_{\mathrm{ov}}\mu_{\mathrm{ev}}^2}\big(3\rho_{\mathrm{ov}} - 6 + \mu_{\mathrm{ev}}\big[12 - 6\rho_{\mathrm{ov}} + 6(1-\rho)\mu_{\mathrm{ov}}\omega_v\big]\omega_{\mathrm{ev}} + 3\lambda_{\mathrm{ev}}'\mu_{\mathrm{ov}}\omega_{\mathrm{ev}}^2\big)$$

then

$$\mathrm{Var}_{\text{Non-Preemp}} < \mathrm{Var}_{\text{Preemp}}$$

where $\mathrm{Var}_{\text{Preemp}}$ and $\mathrm{Var}_{\text{Non-Preemp}}$ are the variances of length of visit under the preemptive-resume and non-preemptive policies, respectively.

Proof: See the Appendix. ■

As one can see, if the third moment $E(S_{\text{ev}}^3)$ is small enough, then $\text{Var}_{\text{Non-Preemp}}$ is smaller than $\text{Var}_{\text{Preemp}}$. Numerical experiments have demonstrated that such a condition is typically met, so that the non-preemptive policy's superiority holds. In particular, when the assumption of the exponentially distributed service and vacation time is introduced, this conclusion is always valid.

Corollary 5: Under assumptions 1)–5), with the exponential service and vacation time distributions, assume $\mu_{\text{ov}} < \mu_{\text{ev}}$, and then

$$\text{Var}_{\text{Non-Preemp}}^{\text{exp}} < \text{Var}_{\text{Preemp}}^{\text{exp}}.$$

Proof: See the Appendix. ■

The comparison of variances of length of visit between the first come first serve and non-preemptive policies is intricate, even in the exponential case. Although a comparison formula can be derived (see the proof of Proposition 9), no simple relationship has been identified. Based on extensive numerical studies, it is observed that the first come first serve policy yields a smaller variance compared with the non-preemptive policy when all the parameters are chosen within the typically range in practice. In particular, under the exponential assumption of service and vacation time distributions, select $(\lambda'_{\text{ov}}/\lambda'_{\text{ev}}) \in (1, 11)$ and $(\mu_{\text{ov}}/\mu_{\text{ev}}) \in (0.1, 1)$. In addition, let $\rho_{\text{ov}} + \rho_{\text{ev}} < 1$. A total of 10 000 examples are randomly generated to evaluate the variance. Without a single exception, in all the experiments carried out, it is always the case that

$$\text{Var}_{\text{FCFS}}^{\text{exp}} \le \text{Var}_{\text{Non-Preemp}}^{\text{exp}}.$$

Therefore, we recommend the first come first serve policy, which generally leads to a smaller mean and variance of patient length of visit.

VIII. Conclusion

In this paper, an analytical model of primary care delivery with e-visits is developed. Formulas to evaluate the mean and variance of lengths of visit for both the office and e-visit patients are derived. Three scheduling policies coordinating the office and the e-visits are compared and the impact of e-visits on primary care delivery is investigated. The model provides a quantitative tool for primary care physicians and clinic managers to apprehend e-visits and design effective care delivery policies. Future work can be directed to the following avenues.

- The current model assumes the Poisson arrivals that may be too strict. Thus, extending the model to general arrivals is essential. Moreover, incorporating the waiting time for office visit appointments and patients' availability or preference, and investigating their impacts are desirable.
- In addition to the three scheduling policies discussed in this paper, other policies, such as blocking (i.e., assign a block time to process e-visits) or alternative (i.e., switching between office and e-visits) are optional. Evaluating and comparing other scheduling policies are also of interest.
- There are many activities conducted during an office visit. In order to understand the impact of e-visits on all the involved activities, detailed care delivery processes within office visits need to be considered.
- In addition, primary care service is usually carried out by a team including physicians, medical assistants, nurses, and so on. Investigating staffing policies in primary care delivery with e-visits is important.
- To improve access, coordination and control of patient flow plays a significant role. Thus, a sensitivity study of arrival rates (λ's) is critical to design the effective control policy to facilitate patient access to care.
- To achieve effective Web services, investigating the scheduling regime and workload of triage nurses who manage the online program is of importance.
- Finally, the results need to be applied on the clinic floor.

Appendix
Proofs

Due to space limitation, we omit the majority of algebraic operations and only provide the sketch of proofs.

A. Proof of Theorem 1

To model physician's office and e-visit services, consider an M/G/1 queue with vacations. Define N_k as the total number of patients in the system after the departure of a patient k, and define M_k as the number of new patients arrived during the service time of the patient k, and then the discrete time process N_k constitutes a Markov chain

$$N_{k+1} = \begin{cases} N_k - 1 + M_{k+1}, & N_k > 0 \\ M_{k+1}, & N_k = 0. \end{cases}$$

Denote

$$\widehat{N}_k = \begin{cases} N_k - 1, & N_k > 0 \\ N_k, & N_k = 0. \end{cases}$$

Then

$$N_{k+1} = \widehat{N}_k + M_{k+1}.$$

Since M and $\widehat{N}$ are independent, the probability generating functions (PGFs) satisfy

$$G_N(z) = G_{\widehat{N}}(z) \cdot G_M(z).$$

To determine $G_{\widehat{N}}(z)$ and $G_M(z)$, we have

$$\begin{aligned} G_{\widehat{N}}(z) &= E[z^{\widehat{N}}] = P\{\widehat{N} = 0\} + \sum_{i=1}^{\infty} z^i P\{\widehat{N} = i\} \\ &= P\{N = 0\}\left(1 - \frac{1}{z}\right) + \frac{1}{z}\sum_{i=0}^{\infty} z^i P\{N = i\} \\ &= \frac{G_N(z) - (1-\rho)(1-z)}{z} \end{aligned}$$

where $\rho = \lambda E[S]$, λ is the arrival rate, and $E[S]$ is the expectation of service time S. According to the Pollaczek–Khinchin transform equation [45, Sec. 5.8],

$$G_M(z) = S^*((1-z)\lambda)$$

where $S^*(s)$ represents the Laplace–Stieltjes transform (LST) of S. Then

$$G_N(z) = \frac{(1-\rho)(1-z)S^*((1-z)\lambda)}{S^*((1-z)\lambda) - z}.$$

Similarly, denote $T^*(s)$ as the LST of the patient cycle time T, and then [45, Sec. 5.8]

$$G_N(z) = T^*((1-z)\lambda)$$
$$T^*(s) = \frac{(1-\rho)sS^*(s)}{s-\lambda+\lambda S^*(s)}.$$

Since the waiting time W and the service time S are independent and $T = W + S$, it is natural that $T^*(s) = W^*(s)S^*(s)$, which implies that

$$W^*(s) = \frac{(1-\rho)s}{s-\lambda+\lambda S^*(s)}.$$

Taking the first and second derivatives of $W^*(s)$ at $s = 0$ and applying L'Hôspital's rule, the mean waiting time $E(W)$, the second moment $E(W^2)$, and the variance of waiting time $Var(W) = E(W^2) - E^2(W)$ can be calculated.

If server vacation is considered, denote the vacation time as V with mean $E(V)$ and LST $V^*(s)$, and the PGF of N is modified as

$$\begin{aligned} G_N(z) &= \frac{(1-\rho)(1-V^*((1-z)\lambda))S^*((1-z)\lambda)}{\lambda E(V)(S^*((1-z)\lambda)-z)} \\ &= T^*((1-z)\lambda). \end{aligned}$$

Through similar derivation, we have

$$W^*(s) = \frac{1-V^*(s)}{sE(V)} \cdot \frac{(1-\rho)s}{s-\lambda+\lambda S^*(s)}.$$

Then, the first and second moments of the waiting time are given by

$$\begin{aligned} E(W) &= \frac{\lambda E(S^2)}{2(1-\rho)} + \frac{E(V^2)}{2E(V)} \\ E(W^2) &= \frac{\lambda E(S^3)}{3(1-\rho)} + \frac{\lambda^2 E(S^2)^2}{2(1-\rho)^2} \\ &\quad + \frac{\lambda E(S^2)E(V^2)}{2(1-\rho)E(V)} + \frac{E(V^3)}{3E(V)}. \end{aligned}$$

In the case of the first come first serve policy, let the total arrival rate be $\lambda = \lambda'_{\text{ov}} + \lambda'_{\text{ev}}$, and define

$$p_i = \lambda'_i/\lambda, \quad i = \text{ev}, \text{ov}.$$

Then, the first to the third moments of the service time S can be evaluated as

$$\begin{aligned} E(S) &= \sum_{i=\text{ev},\text{ov}} p_i \frac{1}{\mu_i} = \left(\frac{\lambda'_{\text{ov}}}{\mu_{\text{ov}}} + \frac{\lambda'_{\text{ev}}}{\mu_{\text{ev}}}\right)\frac{1}{\lambda} = \frac{\rho}{\lambda} \\ E(S^2) &= \sum_{i=\text{ev},\text{ov}} p_i E\left(S_i^2\right) = \frac{2}{\lambda}\left(\frac{\lambda'_{\text{ov}}\delta_{\text{ov}}}{\mu_{\text{ov}}^2} + \frac{\lambda'_{\text{ev}}\delta_{\text{ev}}}{\mu_{\text{ev}}^2}\right) \\ &= \frac{2}{\lambda}(\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}}) \\ E(S^3) &= \sum_{i=\text{ev},\text{ov}} p_i E\left(S_i^3\right) = \frac{\lambda'_{\text{ov}} E\left(S_{\text{ov}}^3\right) + \lambda'_{\text{ev}} E\left(S_{\text{ev}}^3\right)}{\lambda}. \end{aligned}$$

Consequently, the waiting time can be calculated as

$$E(W) = (\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}})\frac{1}{(1-\rho)} + \omega_v$$

and the variance of waiting

$$\begin{aligned} \text{Var}(W) &= E(W^2) - E^2(W) \\ &= \frac{(\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}})^2}{(1-\rho)^2} - \omega_v^2 \\ &\quad + \frac{(1-\rho)\mu_v E(V^3) + \lambda'_{\text{ov}} E\left(S_{\text{ov}}^3\right) + \lambda'_{\text{ev}} E\left(S_{\text{ev}}^3\right)}{3(1-\rho)}. \end{aligned}$$

Finally, the average time each type of patients spend in the system and the associated variance are obtained

$$\begin{aligned} T_i &= E(W) + E(S_i) \\ \text{Var}_i &= \text{Var}(W) + \text{Var}(S_i), \quad i = \text{ev}, \text{ov}. \end{aligned}$$

Next, consider the M/G/1 queue with priorities. The formulations can be derived similarly but with more complexity. The detailed derivation can be referred to [46, Ch. 3.3 and 3.4]. For the non-preemptive policy

$$\begin{aligned} E(W_{\text{ov}}) &= \frac{(1-\rho)\frac{E(V^2)}{E(V)} + \lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right) + \lambda'_{\text{ev}} E\left(S_{\text{ev}}^2\right)}{2(1-\rho_{\text{ov}})} \\ E(W_{\text{ev}}) &= \frac{(1-\rho)\frac{E(V^2)}{E(V)} + \lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right) + \lambda'_{\text{ev}} E\left(S_{\text{ev}}^2\right)}{2(1-\rho_{\text{ov}})(1-\rho)} \\ E(W_{\text{ov}}^2) &= \frac{\lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right)(1-\rho)\frac{E(V^2)}{E(V)}}{2(1-\rho_{\text{ov}})^2} \\ &\quad + \frac{\lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right)\left(\lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right) + \lambda'_{\text{ev}} E\left(S_{\text{ev}}^2\right)\right)}{2(1-\rho_{\text{ov}})^2} \\ &\quad + \frac{(1-\rho)\frac{E(V^3)}{E(V)} + \lambda'_{\text{ov}} E\left(S_{\text{ov}}^3\right) + \lambda'_{\text{ev}} E\left(S_{\text{ev}}^3\right)}{3(1-\rho_{\text{ov}})} \\ E(W_{\text{ev}}^2) &= \frac{(1-\rho)\frac{E(V^2)}{E(V)} + \lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right) + \lambda'_{\text{ev}} E\left(S_{\text{ev}}^2\right)}{2(1-\rho_{\text{ov}})(1-\rho)} \\ &\quad \cdot \left(\frac{\lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right)}{(1-\rho_{\text{ov}})^2} + \frac{\lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right) + \lambda'_{\text{ev}} E\left(S_{\text{ev}}^2\right)}{(1-\rho_{\text{ov}})(1-\rho)}\right) \\ &\quad + \frac{(1-\rho)\frac{E(V^3)}{E(V)} + \lambda'_{\text{ov}} E\left(S_{\text{ov}}^3\right) + \lambda'_{\text{ev}} E\left(S_{\text{ev}}^3\right)}{3(1-\rho_{\text{ov}})^2(1-\rho)}. \end{aligned}$$

The second moments are rephrased as

$$E(S_i^2) = \frac{2\omega_i}{\mu_i}, \quad i = \text{ev}, \text{ov}, v.$$

Plugging in the second moments, the expected patient length of visit and the associated variance for each type of patients can be obtained.

For the preemptive-resume policy

$$\begin{aligned} E(W_{\text{ov}}) &= \frac{(1-\rho)\frac{E(V^2)}{E(V)} + \lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right)}{2(1-\rho_{\text{ov}})} \\ E(W_{\text{ev}}) &= \frac{(1-\rho)\frac{E(V^2)}{E(V)} + \lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right) + \lambda'_{\text{ev}} E\left(S_{\text{ev}}^2\right)}{2(1-\rho_{\text{ov}})(1-\rho)} \\ E(W_{\text{ov}}^2) &= \frac{\lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right)\left((1-\rho)\frac{E(V^2)}{E(V)} + \lambda'_{\text{ov}} E\left(S_{\text{ov}}^2\right)\right)}{2(1-\rho_{\text{ov}})^2} \\ &\quad + \frac{(1-\rho)\frac{E(V^3)}{E(V)} + \lambda'_{\text{ov}} E\left(S_{\text{ov}}^3\right)}{3(1-\rho_{\text{ov}})} \end{aligned}$$

$$E(W_{\text{ev}}^2) = \frac{(1-\rho)\frac{E(V^2)}{E(V)} + \lambda'_{\text{ov}}E\left(S_{\text{ov}}^2\right) + \lambda'_{\text{ev}}E\left(S_{\text{ev}}^2\right)}{2(1-\rho_{\text{ov}})(1-\rho)} \cdot \left(\frac{\lambda'_{\text{ov}}E\left(S_{\text{ov}}^2\right)}{(1-\rho_{\text{ov}})^2} + \frac{\lambda'_{\text{ov}}E\left(S_{\text{ov}}^2\right) + \lambda'_{\text{ev}}E\left(S_{\text{ev}}^2\right)}{(1-\rho_{\text{ov}})(1-\rho)}\right) + \frac{(1-\rho)\frac{E(V^3)}{E(V)} + \lambda'_{\text{ov}}E\left(S_{\text{ov}}^3\right) + \lambda'_{\text{ev}}E\left(S_{\text{ev}}^3\right)}{3(1-\rho_{\text{ov}})^2(1-\rho)}.$$

Plugging in the second moments, the expected patient length of visit and the associated variance for each type of patients can be calculated. ■

B. Proof of Proposition 1

From Theorem 1, take the partial derivative of T_{ov} with respect to β_{ov}

$$\frac{\partial T_{\text{ov}}}{\partial \beta_{\text{ov}}} = \begin{cases} \dfrac{\lambda_{\text{ev}}[\omega_{\text{ov}} + \rho_{\text{ev}}(\omega_{\text{ev}} - \omega_{v})]}{\mu_{\text{ov}}(1-\rho_{\text{ov}})^2}, & \text{non-preemptive policy} \\ \dfrac{\lambda_{\text{ev}}[\omega_{\text{ov}} - \omega_{v}\rho_{\text{ev}}]}{\mu_{\text{ov}}(1-\rho_{\text{ov}})^2}, & \text{preemptive-resume policy} \\ \dfrac{\lambda_{\text{ev}}[\omega_{\text{ov}}(1-\rho_{\text{ev}}) + \omega_{\text{ev}}\rho_{\text{ev}}]}{\mu_{\text{ov}}(1-\rho)^2}, & \text{first come first serve policy.} \end{cases}$$

As one can see, under the first come first serve policy, $(\partial T_{\text{ov}}/\partial \beta_{\text{ov}}) > 0$ without any condition, since $\rho_{\text{ev}} < 1$. Under the non-preemptive policy, $(\partial T_{\text{ov}}/\partial \beta_{\text{ov}}) > 0$ if and only if $\omega_{\text{ov}} + \rho_{\text{ev}}(\omega_{\text{ev}} - \omega_{v}) > 0$. For the preemptive-resume policy, $(\partial T_{\text{ov}}/\partial \beta_{\text{ov}}) > 0$ if and only if $\omega_{\text{ov}} > \omega_{v}\rho_{\text{ev}}$. ■

C. Proof of Proposition 2

From Theorem 1, we obtain

$$\frac{\partial T_{\text{ov}}}{\partial \beta_{\text{ev}}} = \begin{cases} \dfrac{\lambda_{\text{ws}}[\beta_{\text{ov}}\mu_{\text{ev}}\omega_{\text{ov}} - (\lambda'_{\text{ov}} - \beta_{\text{ov}}\lambda_{\text{ev}} - \mu_{\text{ov}})(\omega_{\text{ev}} - \omega_{v})]}{\mu_{\text{ev}}\mu_{\text{ov}}(1-\rho_{\text{ov}})^2}, & \text{non-preemptive policy} \\ \dfrac{\lambda_{\text{ws}}[\beta_{\text{ov}}\mu_{\text{ev}}\omega_{\text{ov}} + (\lambda'_{\text{ov}} - \beta_{\text{ov}}\lambda_{\text{ev}} - \mu_{\text{ov}})\omega_{v}]}{\mu_{\text{ev}}\mu_{\text{ov}}(1-\rho_{\text{ov}})^2}, & \text{preemptive-resume policy} \\ \dfrac{\lambda_{\text{ws}}[\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}(1-\rho_{\text{ov}})]}{\mu_{\text{ev}}(1-\rho)^2} + \dfrac{\lambda_{\text{ws}}\beta_{\text{ov}}[\omega_{\text{ov}}(1-\rho_{\text{ev}}) + \omega_{\text{ev}}\rho_{\text{ev}}]}{\mu_{\text{ov}}(1-\rho)^2}, & \text{first come first serve policy.} \end{cases}$$

Again $(\partial T_{\text{ov}}/\partial \beta_{\text{ev}}) > 0$, unconditionally under the first come first serve policy. After simple algebraic operations, it can be shown that under the non-preemptive policy

$$\frac{\partial T_{\text{ov}}}{\partial \beta_{\text{ev}}} > 0 \text{ iff } \beta_{\text{ov}}\mu_{\text{ev}}\omega_{\text{ov}} > (\mu_{\text{ov}} - \lambda_{\text{ov}})(\omega_{v} - \omega_{\text{ev}}).$$

Under the preemptive-resume policy

$$\frac{\partial T_{\text{ov}}}{\partial \beta_{\text{ev}}} > 0 \quad \text{iff } \beta_{\text{ov}}\mu_{\text{ev}}\omega_{\text{ov}} > (\mu_{\text{ov}} - \lambda_{\text{ov}})\omega_{v}.$$

■

D. Proof of Proposition 3

From Theorem 1, under the non-preemptive policy, we obtain

$$\frac{\partial T_{\text{ev}}}{\partial \beta_{\text{ov}}} = \frac{\lambda_{\text{ev}}(\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}\rho_{\text{ev}} + \omega_{v}(1-\rho))}{\mu_{\text{ov}}(1-\rho_{\text{ov}})(1-\rho)^2} + \frac{\lambda_{\text{ev}}(\omega_{\text{ov}} - \omega_{v})}{\mu_{\text{ov}}(1-\rho_{\text{ov}})(1-\rho)} + \frac{\lambda_{\text{ev}}(\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}\rho_{\text{ev}} + \omega_{v}(1-\rho))}{\mu_{\text{ov}}(1-\rho_{\text{ov}})^2(1-\rho)}$$

$$\frac{\partial T_{\text{ev}}}{\partial \beta_{\text{ev}}} = \frac{\lambda_{\text{ws}}\beta_{\text{ov}}(\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}\rho_{\text{ev}} + \omega_{v}(1-\rho))}{\mu_{\text{ov}}(1-\rho_{\text{ov}})^2(1-\rho)} + \frac{\lambda_{\text{ws}}\left(\frac{\beta_{\text{ov}}\omega_{\text{ov}}}{\mu_{\text{ov}}} + \frac{\omega_{\text{ev}}}{\mu_{\text{ev}}} - \frac{\beta_{\text{ov}}\omega_{v}}{\mu_{\text{ov}}} - \frac{\omega_{v}}{\mu_{\text{ev}}}\right)}{(1-\rho_{\text{ov}})(1-\rho)} + \frac{\lambda_{\text{ws}}\left(\frac{\beta_{\text{ov}}}{\mu_{\text{ov}}} + \frac{1}{\mu_{\text{ev}}}\right)(\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}\rho_{\text{ev}} + \omega_{v}(1-\rho))}{(1-\rho_{\text{ov}})(1-\rho)^2}.$$

Simplifying the above equations, one can show that $(\partial T_{\text{ev}}/\partial \beta_{\text{ov}}) > 0$ if and only if

$$\omega_{\text{ev}}\rho_{\text{ev}}(2-\rho-\rho_{\text{ov}}) + \omega_{\text{ov}}[1-\rho+\rho_{\text{ov}}(1-\rho_{\text{ov}})] + \omega_{v}(1-\rho)^2 > 0$$

which is satisfied without any condition. In addition, $(\partial T_{\text{ev}}/\partial \beta_{\text{ev}}) > 0$ if and only if

$$\rho_{\text{ov}}(1-\rho_{\text{ov}})\omega_{\text{ov}} + (1-\rho_{\text{ov}})^2\omega_{\text{ev}} + \beta_{\text{ov}}\frac{\mu_{\text{ev}}}{\mu_{\text{ov}}}[\omega_{\text{ev}}\rho_{\text{ev}}(2-\rho-\rho_{\text{ov}}) + \omega_{\text{ov}}[1-\rho+\rho_{\text{ov}}(1-\rho_{\text{ov}})] + \omega_{v}(1-\rho)^2] > 0$$

which is always satisfied as well.

Under the preemptive-resume policy, we have

$$\frac{\partial T_{\text{ev}}}{\partial \beta_{\text{ov}}} = \frac{\lambda_{\text{ev}}(\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}\rho_{\text{ev}} + \omega_{v}(1-\rho))}{\mu_{\text{ov}}(1-\rho_{\text{ov}})(1-\rho)^2} + \frac{\lambda_{\text{ev}}(\omega_{\text{ov}} - \omega_{v})}{\mu_{\text{ov}}(1-\rho_{\text{ov}})(1-\rho)} + \frac{\rho_{\text{ev}}}{\mu_{\text{ov}}(1-\rho_{\text{ov}})^2} + \frac{\lambda_{\text{ev}}(\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}\rho_{\text{ev}} + \omega_{v}(1-\rho))}{\mu_{\text{ov}}(1-\rho_{\text{ov}})^2(1-\rho)}$$

$$\frac{\partial T_{\text{ev}}}{\partial \beta_{\text{ev}}} = \frac{\lambda_{\text{ws}}\beta_{\text{ov}}(\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}\rho_{\text{ev}} + \omega_{v}(1-\rho))}{\mu_{\text{ov}}(1-\rho_{\text{ov}})^2(1-\rho)} + \frac{\lambda_{\text{ws}}\left(\frac{\beta_{\text{ov}}\omega_{\text{ov}}}{\mu_{\text{ov}}} + \frac{\omega_{\text{ev}}}{\mu_{\text{ev}}} - \frac{\beta_{\text{ov}}\omega_{v}}{\mu_{\text{ov}}} - \frac{\omega_{v}}{\mu_{\text{ev}}}\right)}{(1-\rho_{\text{ov}})(1-\rho)} + \frac{\lambda_{\text{ws}}\left(\frac{\beta_{\text{ov}}}{\mu_{\text{ov}}} + \frac{1}{\mu_{\text{ev}}}\right)(\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}\rho_{\text{ev}} + \omega_{v}(1-\rho))}{(1-\rho_{\text{ov}})(1-\rho)^2} + \frac{\lambda_{\text{ws}}\beta_{\text{ov}}}{\mu_{\text{ov}}\mu_{\text{ev}}(1-\rho_{\text{ov}})^2}.$$

Comparing with the results under the non-preemptive policy, an additional term $(\rho_{\text{ev}}/\mu_{\text{ov}}(1-\rho_{\text{ov}})^2)$ is added in $(\partial T_{\text{ev}}/\partial \beta_{\text{ov}})$, and an additional term $(\lambda_{\text{ws}}\beta_{\text{ov}}/\mu_{\text{ov}}\mu_{\text{ev}}(1-\rho_{\text{ov}})^2)$ is added in $(\partial T_{\text{ev}}/\partial \beta_{\text{ev}})$. Since both of them are positive, $(\partial T_{\text{ev}}/\partial \beta_{\text{ov}}) > 0$ and $(\partial T_{\text{ev}}/\partial \beta_{\text{ev}}) > 0$ are satisfied.

Under the first come first serve policy, we have

$$\frac{\partial T_{\text{ev}}}{\partial \beta_{\text{ov}}} = \frac{\lambda_{\text{ev}}(\omega_{\text{ov}}(1-\rho_{\text{ev}}) + \omega_{\text{ev}}\rho_{\text{ev}})}{\mu_{\text{ov}}(1-\rho)^2} > 0$$

$$\frac{\partial T_{\text{ev}}}{\partial \beta_{\text{ev}}} = \frac{\lambda_{\text{ws}}(\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}(1-\rho_{\text{ov}}))}{\mu_{\text{ev}}(1-\rho)^2} + \frac{\lambda_{\text{ws}}\beta_{\text{ov}}(\omega_{\text{ov}}(1-\rho_{\text{ev}}) + \omega_{\text{ev}}\rho_{\text{ev}})}{\mu_{\text{ov}}(1-\rho)^2} > 0.$$

■

E. Proof of Proposition 4

First, consider the non-preemptive policy. Define

$$W_{\text{ov}} = \frac{\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}} + (1-\rho)\omega_{\upsilon}}{1-\rho_{\text{ov}}}$$
$$N_{\text{ov}} = \frac{\rho_{\text{ov}}\omega_{\text{ov}} - \rho_{\text{ev}}\omega_{\text{ev}} - (1-\rho)\omega_{\upsilon}}{1-\rho_{\text{ov}}}$$
$$M_{\text{ov}} = \frac{(1-\rho)\mu_{\upsilon}E(V^3) + \lambda'_{\text{ov}}E(S^3_{\text{ov}}) + \lambda'_{\text{ev}}E(S^3_{\text{ev}})}{1-\rho_{\text{ov}}}$$

and

$$\text{Var}(W_{\text{ov}}) = W_{\text{ov}}N_{\text{ov}} + \frac{1}{3}M_{\text{ov}}.$$

The partial derivative of Var_{ov} with respect to β_i can be expressed as

$$\frac{\partial \text{Var}_{\text{ov}}}{\partial \beta_i} = \frac{\partial \text{Var}(W_{\text{ov}})}{\partial \beta_i} = \frac{\partial W_{\text{ov}}}{\partial \beta_i}N_{\text{ov}} + W_{\text{ov}}\frac{\partial N_{\text{ov}}}{\partial \beta_i} + \frac{1}{3}\frac{\partial M_{\text{ov}}}{\partial \beta_i}, \quad i = \text{ev}, \text{ov}.$$

It can be shown that

$$\frac{\partial W_{\text{ov}}}{\partial \beta_{\text{ov}}} > 0 \text{ iff } \omega_{\text{ov}} + \omega_{\text{ev}}\rho_{\text{ev}} - \omega_{\upsilon}\rho_{\text{ev}} > 0$$
$$\frac{\partial N_{\text{ov}}}{\partial \beta_{\text{ov}}} > 0 \text{ iff } \omega_{\text{ov}} - \omega_{\text{ev}}\rho_{\text{ev}} + \omega_{\upsilon}\rho_{\text{ev}} > 0$$
$$\frac{\partial M_{\text{ov}}}{\partial \beta_{\text{ov}}} > 0 \text{ iff } \mu_{\text{ov}}E(S^3_{\text{ov}}) + \lambda'_{\text{ev}}E(S^3_{\text{ev}}) - \rho_{\text{ev}}\mu_{\upsilon}E(V^3) > 0.$$

The sufficient conditions are satisfied when all the above-mentioned inequalities are true.

Next, consider the preemptive-resume policy. Define

$$W_{\text{ov}} = \frac{\rho_{\text{ov}}\omega_{\text{ov}} + (1-\rho)\omega_{\upsilon}}{1-\rho_{\text{ov}}}$$
$$N_{\text{ov}} = \frac{\rho_{\text{ov}}\omega_{\text{ov}} - (1-\rho)\omega_{\upsilon}}{1-\rho_{\text{ov}}}$$
$$M_{\text{ov}} = \frac{(1-\rho)\mu_{\upsilon}E(V^3) + \lambda'_{\text{ov}}E(S^3_{\text{ov}})}{1-\rho_{\text{ov}}}.$$

Similarly, one can show that $(\partial W_{\text{ov}}/\partial \beta_{\text{ov}}) > 0$ implies $\omega_{\text{ov}} > \omega_{\upsilon}\rho_{\text{ev}}$, $(\partial M_{\text{ov}}/\partial \beta_{\text{ov}}) > 0$ implies $\mu_{\text{ov}}E(S^3_{\text{ov}}) > \rho_{\text{ev}}\mu_{\upsilon}E(S^3_{\upsilon})$, and $(\partial N_{\text{ov}}/\partial \beta_{\text{ov}}) > 0$ is always true, since $\omega_{\text{ov}} + \omega_{\upsilon}\rho_{\text{ev}} > 0$.

Finally, consider the first come first serve policy. Define

$$W = \frac{\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}} + (1-\rho)\omega_{\upsilon}}{1-\rho}$$
$$N = \frac{\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}} - (1-\rho)\omega_{\upsilon}}{1-\rho}$$
$$M = \frac{(1-\rho)\mu_{\upsilon}E(S^3_{\upsilon}) + \lambda'_{\text{ov}}E(S^3_{\text{ov}}) + \lambda'_{\text{ev}}E(S^3_{\text{ev}})}{1-\rho}.$$

It can be shown that both $(\partial W/\partial \beta_{\text{ov}}) > 0$ and $(\partial N/\partial \beta_{\text{ov}}) > 0$ imply $\omega_{\text{ov}}(1-\rho_{\text{ev}}) + \omega_{\text{ev}}\rho_{\text{ev}} > 0$, and $(\partial M/\partial \beta_{\text{ov}}) > 0$ implies $\mu_{\text{ov}}(1-\rho_{\text{ev}})E(S^3_{\text{ov}}) + \lambda'_{\text{ev}}E(S^3_{\text{ev}}) > 0$, which are always true. ■

F. Proof of Proposition 5

First, consider the non-preemptive policy

$$\frac{\partial W_{\text{ov}}}{\partial \beta_{\text{ev}}} > 0 \text{ iff } \beta_{\text{ov}}\omega_{\text{ov}} > (\omega_{\upsilon} - \omega_{\text{ev}})\left[\beta_{\text{ov}}\rho_{\text{ev}} + \frac{\mu_{\text{ov}}}{\mu_{\text{ev}}}(1-\rho_{\text{ov}})\right]$$
$$\frac{\partial N_{\text{ov}}}{\partial \beta_{\text{ev}}} > 0 \text{ iff } \beta_{\text{ov}}\omega_{\text{ov}} > (\omega_{\text{ev}} - \omega_{\upsilon})\left[\beta_{\text{ov}}\rho_{\text{ev}} + \frac{\mu_{\text{ov}}}{\mu_{\text{ev}}}(1-\rho_{\text{ov}})\right]$$
$$\frac{\partial M_{\text{ov}}}{\partial \beta_{\text{ev}}} > 0 \text{ iff } \beta_{\text{ov}}\left[\mu_{\text{ov}}E(S^3_{\text{ov}}) + \lambda'_{\text{ev}}E(S^3_{\text{ev}}) - \rho_{\text{ev}}\mu_{\upsilon}E(S^3_{\upsilon})\right] + (1-\rho_{\text{ov}})\frac{\mu_{\text{ov}}}{\mu_{\text{ev}}}\left[\mu_{\text{ev}}E(S^3_{\text{ev}}) - \mu_{\upsilon}E(S^3_{\upsilon})\right] > 0.$$

Plug- in $\rho_{\text{ev}} = (\lambda_{\text{ev}} + \beta_{\text{ev}}\lambda_{\text{ws}})/\mu_{\text{ev}}$ and $\rho_{\text{ov}} = (\lambda_{\text{ov}} + \beta_{\text{ov}}(\lambda_{\text{ev}} + \beta_{\text{ev}}\lambda_{\text{ws}}))/\mu_{\text{ov}}$ Then, the sufficient but not necessary conditions for $(\partial \text{Var}_{\text{ov}}/\partial \beta_{\text{ev}}) > 0$ are $\beta_{\text{ov}}\mu_{\text{ev}}\omega_{\text{ov}} > (\mu_{\text{ov}} - \lambda_{\text{ov}})|\omega_{\upsilon} - \omega_{\text{ev}}|$ and $\mu_{\text{ev}}E(S^3_{\text{ev}}) \geq \mu_{\upsilon}E(S^3_{\upsilon})$.

Next, consider the preemptive-resume policy

$$\frac{\partial W_{\text{ov}}}{\partial \beta_{\text{ev}}} > 0 \text{ iff } \beta_{\text{ov}}(\omega_{\text{ov}} - \omega_{\upsilon}\rho_{\text{ev}}) - (1-\rho_{\text{ov}})\omega_{\upsilon}\frac{\mu_{\text{ov}}}{\mu_{\text{ev}}} > 0$$
$$\frac{\partial N_{\text{ov}}}{\partial \beta_{\text{ev}}} > 0 \text{ iff } \beta_{\text{ov}}(\omega_{\text{ov}} + \omega_{\upsilon}\rho_{\text{ev}}) + (1-\rho_{\text{ov}})\omega_{\upsilon}\frac{\mu_{\text{ov}}}{\mu_{\text{ev}}} > 0$$
$$\frac{\partial M_{\text{ov}}}{\partial \beta_{\text{ev}}} > 0 \text{ iff } \beta_{\text{ov}}\left[\mu_{\text{ov}}E(S^3_{\text{ov}}) - \rho_{\text{ev}}\mu_{\upsilon}E(S^3_{\upsilon})\right] - (1-\rho_{\text{ov}})\frac{\mu_{\text{ov}}}{\mu_{\text{ev}}}\mu_{\upsilon}E(S^3_{\upsilon}) > 0.$$

The sufficient but not necessary conditions are $(\partial W_{\text{ov}}/\partial \beta_{\text{ev}}) \geq 0$ and $(\partial M_{\text{ov}}/\partial \beta_{\text{ev}}) \geq 0$, which lead to $\beta_{\text{ov}}\mu_{\text{ev}}\omega_{\text{ov}} > (\mu_{\text{ov}} - \lambda_{\text{ov}})\omega_{\upsilon}$ and $\beta_{\text{ov}}\mu_{\text{ov}}\mu_{\text{ev}}E(S^3_{\text{ov}}) \geq (\mu_{\text{ov}} - \lambda_{\text{ov}})\mu_{\upsilon}E(S^3_{\upsilon})$.

Finally, consider the first come first serve policy

$$\frac{\partial W}{\partial \beta_{\text{ev}}} > 0 \text{ iff } \mu_{\text{ov}}(\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}(1-\rho_{\text{ov}})) + \beta_{\text{ov}}\mu_{\text{ev}}[\omega_{\text{ov}}(1-\rho_{\text{ev}}) + \omega_{\text{ev}}\rho_{\text{ev}}] > 0$$
$$\frac{\partial N}{\partial \beta_{\text{ev}}} > 0 \text{ iff } \beta_{\text{ov}}(\omega_{\text{ov}}(1-\rho_{\text{ev}}) + \omega_{\text{ev}}\rho_{\text{ev}}) + \frac{\mu_{\text{ov}}}{\mu_{\text{ev}}}[\omega_{\text{ov}}\rho_{\text{ov}} + \omega_{\text{ev}}(1-\rho_{\text{ov}})] > 0$$
$$\frac{\partial M}{\partial \beta_{\text{ev}}} > 0 \text{ iff } \lambda_{\text{ov}}\mu_{\text{ov}}E(S^3_{\text{ov}}) + \mu_{\text{ov}}\mu_{\text{ev}}(1-\rho_{\text{ov}})E(S^3_{\text{ev}}) + \mu_{\text{ev}}\beta_{\text{ov}}\left[\mu_{\text{ov}}(1-\rho_{\text{ev}})E(S^3_{\text{ov}}) + \lambda_{\text{ev}}E(S^3_{\text{ev}})\right] > 0$$

which are always true. ■

G. Proof of Proposition 6

Consider both the non-preemptive and preemptive-resume policies. Define

$$W_{\text{ev}} = \frac{\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}} + (1-\rho)\omega_{\upsilon}}{(1-\rho_{\text{ov}})(1-\rho)}$$
$$N_{\text{ev}} = \frac{\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}} - (1-\rho)\omega_{\upsilon}}{(1-\rho_{\text{ov}})(1-\rho)} + \frac{2\rho_{\text{ov}}\omega_{\text{ov}}}{(1-\rho_{\text{ov}})^2}$$
$$M_{\text{ev}} = \frac{(1-\rho)\mu_{\upsilon}E(S^3_{\upsilon}) + \lambda'_{\text{ov}}E(S^3_{\text{ov}}) + \lambda'_{\text{ev}}E(S^3_{\text{ev}})}{(1-\rho_{\text{ov}})^2(1-\rho)}.$$

Under the non-preemptive policy

$$\text{Var}(W_{\text{ev}}) = W_{\text{ev}}N_{\text{ev}} + \frac{1}{3}M_{\text{ev}}$$

$$\frac{\partial \mathrm{Var}_{\mathrm{ev}}}{\partial \beta_i} = \frac{\partial \mathrm{Var}(W_{\mathrm{ev}})}{\partial \beta_i} = \frac{\partial W_{\mathrm{ev}}}{\partial \beta_i} N_{\mathrm{ev}} + W_{\mathrm{ev}} \frac{\partial N_{\mathrm{ev}}}{\partial \beta_i} + \frac{1}{3}\frac{\partial M_{\mathrm{ev}}}{\partial \beta_i}, \quad i = \mathrm{ev}, \mathrm{ov}.$$

In addition, under the preemptive-resume policy

$$A_{\mathrm{ev}} = \frac{2\delta_{\mathrm{ev}} - 1}{\mu_{\mathrm{ev}}^2 (1-\rho_{\mathrm{ov}})^2}$$

$$\frac{\partial \mathrm{Var}_{\mathrm{ev}}}{\partial \beta_i} = \frac{\partial \mathrm{Var}(W_{\mathrm{ev}})}{\partial \beta_i} + \frac{\partial A_{\mathrm{ev}}}{\partial \beta_i}, \quad i = \mathrm{ev}, \mathrm{ov}.$$

It follows that:

$$\begin{aligned}
\frac{\partial W_{\mathrm{ev}}}{\partial \beta_{\mathrm{ov}}} > 0 \text{ iff } & \omega_{\mathrm{ov}}\big(1-\rho_{\mathrm{ov}}^2-\rho_{\mathrm{ev}}\big) \\
& + \omega_{\mathrm{ev}}\rho_{\mathrm{ev}}(2-\rho_{\mathrm{ov}}-\rho) + \omega_v(1-\rho)^2 > 0 \\
\frac{\partial N_{\mathrm{ev}}}{\partial \beta_{\mathrm{ov}}} > 0 \text{ iff } & \omega_{\mathrm{ov}}\big(\big(1-\rho_{\mathrm{ov}}^2-\rho_{\mathrm{ev}}\big)(1-\rho_{\mathrm{ov}}) \\
& + 2(1+\rho_{\mathrm{ov}})(1-\rho)^2\big) \\
& + \omega_{\mathrm{ev}}\rho_{\mathrm{ev}}(2-\rho_{\mathrm{ov}}-\rho)(1-\rho_{\mathrm{ov}}) \\
& - \omega_v(1-\rho)^2(1-\rho_{\mathrm{ov}}) > 0 \\
\frac{\partial M_{\mathrm{ev}}}{\partial \beta_{\mathrm{ov}}} > 0 \text{ iff } & \mu_{\mathrm{ov}}\big[(1+\rho_{\mathrm{ov}})(1-\rho_{\mathrm{ev}}) - 2\rho_{\mathrm{ov}}^2\big]E\big(S_{\mathrm{ov}}^3\big) \\
& + \lambda_{\mathrm{ev}}(3-2\rho-\rho_{\mathrm{ov}})E\big(S_{\mathrm{ev}}^3\big) \\
& + 2\mu_v(1-\rho)^2 E\big(S_v^3\big) > 0.
\end{aligned}$$

In addition

$$\frac{\partial A_{\mathrm{ev}}}{\partial \beta_{\mathrm{ov}}} = \frac{2\rho_{\mathrm{ev}}(2\delta_{\mathrm{ev}}-1)}{\mu_{\mathrm{ov}}\mu_{\mathrm{ev}}(1-\rho_{\mathrm{ov}})^3} > 0.$$

Since all the other conditions are true except for $(\partial N_{\mathrm{ev}}/\partial \beta_{\mathrm{ov}}) > 0$, a sufficient but not necessary condition for $(\partial \mathrm{Var}_{\mathrm{ov}}/\partial \beta_{\mathrm{ov}}) > 0$ is $(\partial N_{\mathrm{ev}}/\partial \beta_{\mathrm{ov}}) \geq 0$. Such a condition can be rephrased as

$$\begin{aligned}
&\omega_{\mathrm{ov}}((1-\rho)(1-\rho_{\mathrm{ov}}) + 2(1-\rho_{\mathrm{ov}})(1-\rho)^2) \\
&\quad + 2\omega_{\mathrm{ev}}\rho_{\mathrm{ev}}(1-\rho)(1-\rho_{\mathrm{ov}}) - \omega_v(1-\rho)^2(1-\rho_{\mathrm{ov}}) \geq 0
\end{aligned}$$

which again can be further relaxed to

$$2\omega_{\mathrm{ov}} - \omega_v \geq 0.$$

Moreover, for β_{ev}

$$\begin{aligned}
\frac{\partial W_{\mathrm{ev}}}{\partial \beta_{\mathrm{ev}}} > 0 \text{ iff } & \frac{\mu_{\mathrm{ov}}}{\mu_{\mathrm{ev}}}(1-\rho_{\mathrm{ov}})[\rho_{\mathrm{ov}}\omega_{\mathrm{ov}} + (1-\rho_{\mathrm{ov}})\omega_{\mathrm{ev}}] \\
& + \beta_{\mathrm{ov}}\big[\omega_{\mathrm{ov}}\big(1-\rho_{\mathrm{ov}}^2-\rho_{\mathrm{ev}}\big) \\
& + \omega_{\mathrm{ev}}\rho_{\mathrm{ev}}(2-\rho_{\mathrm{ov}}-\rho) + \omega_v(1-\rho)^2\big] > 0 \\
\frac{\partial N_{\mathrm{ev}}}{\partial \beta_{\mathrm{ev}}} > 0 \text{ iff } & \frac{\mu_{\mathrm{ov}}}{\mu_{\mathrm{ev}}}(1-\rho_{\mathrm{ov}})^2[\rho_{\mathrm{ov}}\omega_{\mathrm{ov}} + (1-\rho_{\mathrm{ov}})\omega_{\mathrm{ev}}] \\
& + \beta_{\mathrm{ov}}\big[\omega_{\mathrm{ov}}\big(\big(1-\rho_{\mathrm{ov}}^2-\rho_{\mathrm{ev}}\big)(1-\rho_{\mathrm{ov}}) \\
& + 2(1+\rho_{\mathrm{ov}})(1-\rho)^2\big) \\
& + \omega_{\mathrm{ev}}\rho_{\mathrm{ev}}(2-\rho_{\mathrm{ov}}-\rho)(1-\rho_{\mathrm{ov}}) \\
& - \omega_v(1-\rho)^2(1-\rho_{\mathrm{ov}})\big] > 0 \\
\frac{\partial M_{\mathrm{ev}}}{\partial \beta_{\mathrm{ev}}} > 0 \text{ iff } & \mu_{\mathrm{ov}}\big[\lambda_{\mathrm{ov}}(1-\rho_{\mathrm{ov}}) \\
& + \beta_{\mathrm{ov}}\mu_{\mathrm{ev}}\big((1+\rho_{\mathrm{ov}})(1-\rho_{\mathrm{ev}}) - 2\rho_{\mathrm{ov}}^2\big)\big]E\big(S_{\mathrm{ov}}^3\big) \\
& + \mu_{\mathrm{ev}}\big[\mu_{\mathrm{ov}}(1-\rho_{\mathrm{ov}})^2 \\
& + \beta_{\mathrm{ov}}\lambda_{\mathrm{ev}}(3-2\rho-\rho_{\mathrm{ov}})\big]E\big(S_{\mathrm{ev}}^3\big) \\
& + 2\beta_{\mathrm{ov}}\mu_{\mathrm{ev}}\mu_v(1-\rho)^2E\big(S_v^3\big) > 0.
\end{aligned}$$

In addition, in the case of the preemptive-resume policy

$$\frac{\partial A_{\mathrm{ev}}}{\partial \beta_{\mathrm{ev}}} = \frac{2\beta_{\mathrm{ov}}\lambda_{\mathrm{ws}}(2\delta_{\mathrm{ev}}-1)}{\mu_{\mathrm{ov}}\mu_{\mathrm{ev}}^2(1-\rho_{\mathrm{ov}})^3} > 0.$$

Again all the other conditions are true except for $(\partial N_{\mathrm{ev}}/\partial \beta_{\mathrm{ev}}) > 0$, and a sufficient but not necessary condition for $(\partial \mathrm{Var}_{\mathrm{ov}}/\partial \beta_{\mathrm{ev}}) > 0$ is $(\partial N_{\mathrm{ev}}/\partial \beta_{\mathrm{ev}}) \geq 0$. Note that such a condition is the same as the one for $(\partial N_{\mathrm{ev}}/\partial \beta_{\mathrm{ov}}) \geq 0$ by adding one more positive term. Thus, the same sufficient condition applies. ■

H. Proof of Proposition 7

In the case of the first come first serve policy, since the waiting time for both the types of patients are the same, using the results from Propositions 4 and 5, $(\partial \mathrm{Var}_{\mathrm{ev}}/\partial \beta_{\mathrm{ov}}) > 0$ and $(\partial \mathrm{Var}_{\mathrm{ev}}/\partial \beta_{\mathrm{ev}}) > 0$ hold for any choice of parameters. ■

I. Proof of Proposition 8

First, compare $T_{\text{Non-Preemp}}$ and T_{Preemp}

$$T_{\text{Non-Preemp}} - T_{\text{Preemp}} = \frac{\rho_{\mathrm{ov}}\rho_{\mathrm{ev}}\big(\frac{\mu_{\mathrm{ov}}}{\mu_{\mathrm{ev}}}\delta_{\mathrm{ev}} - 1\big)}{(1-\rho_{\mathrm{ov}})\lambda}.$$

Since $\mu_{\mathrm{ov}} < \mu_{\mathrm{ev}}$ and $\mathrm{cv}_{\mathrm{ev}} < 1$, we obtain $T_{\text{Non-Preemp}} < T_{\text{Preemp}}$. Thus, the non-preemptive service system is preferred compared with the preemptive-resume one.

Next, compare non-preemptive and first come first serve policies

$$T_{\text{Non-Preemp}} - T_{\text{FCFS}} = \frac{\rho_{\mathrm{ov}}\omega_{\mathrm{ov}} + \rho_{\mathrm{ev}}\omega_{\mathrm{ev}} + (1-\rho)\omega_v}{(1-\rho_{\mathrm{ov}})(1-\rho)\lambda} \cdot \lambda'_{\mathrm{ov}}\lambda'_{\mathrm{ev}}\left(\frac{1}{\mu_{\mathrm{ov}}} - \frac{1}{\mu_{\mathrm{ev}}}\right).$$

Since $\mu_{\mathrm{ov}} < \mu_{\mathrm{ev}}$, we obtain $T_{\text{Non-Preemp}} > T_{\text{FCFS}}$. Thus, the first come first serve policy yields the shortest overall patient length of visit. ■

J. Proof of Proposition 9

For the variance of the non-preemptive and preemptive-resume policies

$$\begin{aligned}
&\mathrm{Var}_{\text{Non-Preemp}} - \mathrm{Var}_{\text{Preemp}} \\
&\quad = p_{\mathrm{ov}}\left(\frac{\lambda_{\mathrm{ev}}E\big(S_{\mathrm{ev}}^3\big)}{3(1-\rho_{\mathrm{ov}})} - \frac{\rho_{\mathrm{ev}}\omega_{\mathrm{ev}}(\rho_{\mathrm{ev}}\omega_{\mathrm{ev}} + 2(1-\rho)\omega_v)}{(1-\rho_{\mathrm{ov}})^2}\right) \\
&\quad\quad + p_{\mathrm{ev}}\left(1 - \frac{1}{(1-\rho_{\mathrm{ov}})^2}\right)\left(\frac{2\delta_{\mathrm{ev}}-1}{\mu_{\mathrm{ev}}^2}\right).
\end{aligned}$$

It follows that $\mathrm{Var}_{\text{Non-Preemp}} - \mathrm{Var}_{\text{Preemp}} < 0$ if and only if

$$\begin{aligned}
&3\rho_{\mathrm{ov}} - 6 + \mu_{\mathrm{ev}}[12 - 6\rho_{\mathrm{ov}} + 6(1-\rho)\mu_{\mathrm{ov}}\omega_v]\omega_{\mathrm{ev}} \\
&\quad + 3\lambda'_{\mathrm{ev}}\mu_{\mathrm{ov}}\omega_{\mathrm{ev}}^2 - (1-\rho_{\mathrm{ov}})\mu_{\mathrm{ov}}\mu_{\mathrm{ev}}^2E\big(S_{\mathrm{ev}}^3\big) > 0.
\end{aligned}$$

Here, we consider the third moment of the e-visit service, and the condition becomes

$$\begin{aligned}
E\big(S_{\mathrm{ev}}^3\big) < & \frac{1}{(1-\rho_{\mathrm{ov}})\mu_{\mathrm{ov}}\mu_{\mathrm{ev}}^2}\big(3\rho_{\mathrm{ov}} - 6 \\
& + \mu_{\mathrm{ev}}[12 - 6\rho_{\mathrm{ov}} + 6(1-\rho)\mu_{\mathrm{ov}}\omega_v]\omega_{\mathrm{ev}} + 3\lambda'_{\mathrm{ev}}\mu_{\mathrm{ov}}\omega_{\mathrm{ev}}^2\big).
\end{aligned}$$

■

K. Proof of Corollary 5

When all the service and vacation time distributions are exponential and $\mu_{\text{ov}} < \mu_{\text{ev}}$

$$\begin{aligned}\text{Var}^{\text{exp}}_{\text{Non-Preemp}} - \text{Var}^{\text{exp}}_{\text{Preemp}} < 0 \text{ iff } &2(1-\rho)\lambda'_{\text{ov}}\mu_{\text{ev}} \\ &+ \left[(2-\rho_{\text{ov}})\rho_{\text{ov}}\mu_{\text{ev}} - \lambda'_{\text{ov}}(2-\rho_{\text{ov}}-\rho)\right]\mu_v > 0.\end{aligned}$$

Let $\lambda'_{\text{ov}} = a\lambda'_{\text{ev}}$, $\mu_{\text{ov}} = b\mu_{\text{ev}}$, and $b < 1$

$$\begin{aligned}&(2-\rho_{\text{ov}})\rho_{\text{ov}}\mu_{\text{ev}} - \lambda'_{\text{ov}}(1-\rho_{\text{ov}}+1-\rho) > 0 \text{ iff} \\ &\qquad\left(-2a + \frac{2a}{b}\right) + \left(a - \frac{a^2}{b^2} + \frac{2a^2}{b}\right)\rho_{\text{ev}} > 0.\end{aligned}$$

If $(a-(a^2/b^2)+(2a^2/b)) \geq 0$, the above equation is satisfied. If not, the above inequality can be written as

$$\left(2a - \frac{2a}{b}\right) + \left(-a + \frac{a^2}{b^2} - \frac{2a^2}{b}\right)\rho_{\text{ev}} < 0.$$

Since $\rho < 1$, which implies $1-(1+(a/b))\rho_{\text{ev}} > 0$, it leads to $\rho_{ev} < b/(a+b)$. Then, we obtain

$$\begin{aligned}&\left(2a - \frac{2a}{b}\right) + \left(-a + \frac{a^2}{b^2} - \frac{2a^2}{b}\right)\rho_{\text{ev}} \\ &\quad < 2\left(a - \frac{a}{b}\right) + \left(-a + \frac{a^2}{b^2} - \frac{2a^2}{b}\right)\frac{b}{a+b} \\ &\quad = -\frac{a[a+(2-b)b]}{b(a+b)} < 0.\end{aligned}$$

Next, compare Var_{FCFS} and $\text{Var}_{\text{Non-Preemp}}$. Define

$$\begin{aligned}M &= (1-\rho)\mu_v E(V^3) + \lambda'_{\text{ov}} E\left(S^3_{\text{ov}}\right) + \lambda'_{\text{ev}} E\left(S^3_{\text{ev}}\right) \\ N &= \rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}} + (1-\rho)\omega_v.\end{aligned}$$

Then the difference in variance between the two policies is

$$\begin{aligned}&\text{Var}^{\text{exp}}_{\text{Non-Preemp}} - \text{Var}^{\text{exp}}_{\text{FCFS}} \\ &= \left(\frac{p_{\text{ov}}}{1-\rho_{\text{ov}}} + \frac{p_{\text{ev}}}{(1-\rho_{\text{ov}})^2(1-\rho)} - \frac{1}{1-\rho}\right)\frac{M}{3} \\ &\quad + \left[p_{\text{ov}}\frac{\rho_{\text{ov}}\omega_{\text{ov}} - \rho_{\text{ev}}\omega_{\text{ev}} - (1-\rho)\omega_v}{(1-\rho_{\text{ov}})^2}\right. \\ &\quad + p_{\text{ev}}\left(\frac{\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}} - (1-\rho)\omega_v}{(1-\rho_{\text{ov}})^2(1-\rho)^2} + \frac{2\rho_{\text{ov}}\omega_{\text{ov}}}{(1-\rho_{\text{ov}})^3(1-\rho)}\right) \\ &\quad \left. - \frac{\rho_{\text{ov}}\omega_{\text{ov}} + \rho_{\text{ev}}\omega_{\text{ev}} - (1-\rho)\omega_v}{(1-\rho)^2}\right] N.\end{aligned}$$

■

REFERENCES

[1] *The Impending Collapse of Primary Care Medicine and Its Implications for the State of the Nation's Health Care*, Amer. College Phys., Washington, DC, USA, 2006.

[2] T. Bodenheimer, "Primary care—Will it survive?" *New England J. Med.*, vol. 355, no. 9, pp. 861–864, 2006.

[3] *Access is the Answer: Community Health Centers, Primary Care & the Future of American Health Care*, Nat. Assoc. Commun. Health Centers, Bethesda, MD, USA, Mar. 2014.

[4] C. E. Erikson, S. Danish, K. C. Jones, S. F. Sandberg, and A. C. Carle, "The role of medical school culture in primary care career choice," *Acad. Med.*, vol. 88, no. 12, pp. 1919–1926, 2013.

[5] N. Gidwani, L. Fernandez, and D. Schlossman, "Connecting with patients online: E-visits," U.S. Dept. Family Commun. Med. Academic Health Center, Tech. Rep., 2012. [Online]. Available: http://www.medinfodoc.net/med-inf-498-capstone-paper.html

[6] S. C. Adamson and J. W. Bachman, "Pilot study of providing online care in a primary care setting," *Mayo Clin. Proc.*, vol. 85, no. 8, pp. 704–710, 2010.

[7] D. Baer, "Patient-physician e-mail communication: The Kaiser Permanente experience," *J. Oncol. Pract.*, vol. 7, no. 4, pp. 230–233, 2011.

[8] A. Mehrotra *et al.*, "Comparing costs and quality of care at retail clinics with that of other medical settings for 3 common illnesses," *Ann. Internal Med.*, vol. 151, no. 5, pp. 321–328, 2012.

[9] A. J. Watson, H. Bergman, C. M. William, and J. C. Kvedar, "A randomized trial to evaluate the efficacy of online follow-up visits in the management of acne," *Arch. Dermatol.*, vol. 146, no. 4, pp. 406–411, 2010.

[10] A. Mehrotra, S. Paone, G. D. Martich, S. M. Albert, and G. J. Shevchik, "A comparison of care at E-visits and physician office visits for sinusitis and urinary tract infection," *JAMA Internal Med.*, vol. 173, no. 1, pp. 72–74, 2012.

[11] P. Whitten, L. Buis, and B. Love, "Physician-patient e-visit programs: Implementation and appropriateness," *Disease Manage. Health Outcomes*, vol. 15, no. 4, pp. 207–214, 2007.

[12] H. Bavafa, S. Savin, and C. Terwiesch, "Managing office revisit intervals and patient panel sizes in primary care," *Social Sci. Res. Netw.*, 2013. [Online]. Available: http://ssrn.com/abstract=2363685

[13] P. Bower, S. Campbell, C. Bojke, and B. Sibbald, "Team structure, team climate and the quality of care in primary care: An observational study," *Quality Safety Health Care*, vol. 12, no. 4, pp. 273–279, 2003.

[14] L. Lemieux-Charles and W. L. McGuire, "What do we know about health care team effectiveness? A review of the literature," *Med. Care Res. Rev.*, vol. 63, no. 3, pp. 263–300, 2006.

[15] R. Hillestad *et al.*, "Can electronic medical record systems transform health care? Potential health benefits, savings, and costs," *Health Affairs*, vol. 24, no. 5, pp. 1103–1117, 2005.

[16] J. W. Beasley *et al.*, "Information chaos in primary care: Implications for physician performance and patient safety," *J. Amer. Board Family Med.*, vol. 24, no. 6, pp. 745–751, 2011.

[17] T. C. Rosenthal, "The medical home: Growing evidence to support a new approach to primary care," *J. Amer. Board Family Med.*, vol. 21, no. 5, pp. 427–440, 2008.

[18] B. E. Landon, J. M. Gill, R. C. Antonelli, and E. C. Rich, "Prospects for rebuilding primary care using the patient-centered medical home," *Health Affairs*, vol. 29, no. 5, pp. 827–834, 2010.

[19] A. H. Goroll, R. A. Berenson, S. C. Schoenbaum, and L. B. Gardner, "Fundamental reform of payment for adult primary care: Comprehensive payment for comprehensive care," *J. General Internal Med.*, vol. 22, no. 3, pp. 410–415, 2007.

[20] M. E. Porter, E. A. Pabo, and T. H. Lee, "Redesigning primary care: A strategic vision to improve value by organizing around patients' needs," *Health Affairs*, vol. 32, pp. 516–525, Mar. 2013.

[21] C. Salisbury, "Does advanced access work for patients and practices?" *Brit. J. General Pract.*, vol. 54, no. 502, pp. 330–331, 2004.

[22] M. Pickin, A. O'Cathain, F. C. Sampson, and S. Dixon, "Evaluation of advanced access in the national primary care collaborative," *Brit. J. General Pract.*, vol. 54, no. 502, pp. 334–340, 2004.

[23] T. Bodenheimer and H. H. Pham, "Primary care: Current problems and proposed solutions," *Health Affairs*, vol. 29, no. 5, pp. 799–805, 2010.

[24] C. A. Sinsky, R. Willard-Grace, A. M. Schutzbank, T. A. Sinsky, D. Margolius, and T. Bodenheimer, "In search of joy in practice: A report of 23 high-functioning primary care practices," *Ann. Family Med.*, vol. 11, no. 3, pp. 272–278, 2013.

[25] X. Zhong, M. Williams, J. Li, S. Kraft, and J. Sleeth, "Primary care redesign: Review and a simulation study at a pediatric clinic," in *Healthcare Analytics: From Data to Knowledge to Healthcare Improvement* (Wiley Series on Operations Research and Management Science (WORMS)), H. Yang and E. K. Lee, Eds. New York, NY, USA: Wiley, 2016, pp. 399–426.

[26] H. Bavafa, L. M. Hitt, and C. Terwiesch, "Patient portals in primary care: Impacts on patient health and physician productivity," *Social Sci. Res. Netw.*, 2013. [Online]. Available: http://ssrn.com/abstract=2363705

[27] S. H. Jacobson, S. N. Hall, and J. R. Swisher, "Discrete-event simulation of health care systems," in *Patient Flow: Reducing Delay in Healthcare Delivery* (International Series in Operations Research & Management Science), vol. 91. New York, NY, USA: Springer, 2006, pp. 211–252.

[28] T. Eldabi, R. J. Paul, and T. Young, "Simulation modelling in healthcare: Reviewing legacies and investigating Futures," *J. Oper. Res. Soc.*, vol. 58, no. 2, pp. 262–270, 2007.

[29] M. M. Günal and M. Pidd, "Discrete event simulation for performance modelling in health care: A review of the literature," *J. Simul.*, vol. 4, no. 1, pp. 42–51, 2010.

[30] J. L. Wiler, R. T. Griffey, and T. Olsen, "Review of modeling approaches for emergency department patient flow and crowding research," *Acad. Emergency Med.*, vol. 18, no. 12, pp. 1371–1379, 2011.

[31] D. Gupta and B. Denton, "Appointment scheduling in health care: Challenges and opportunities," *IIE Trans.*, vol. 40, no. 9, pp. 800–819, 2008.

[32] L. Green, "Queueing analysis in healthcare," in *Patient Flow: Reducing Delay in Healthcare Delivery*, vol. 91, R. W. Hall, Ed. Springer, 2006, pp. 281–307.

[33] S. Fomundam and J. W. Herrmann, "A survey of queueing theory applications in healthcare," Inst. Syst. Res., Univ. Maryland, College Park, College Park, MD, USA, ISR Tech. Rep. 2007-24, 2007.

[34] L. Jiang and R. E. Giachetti, "A queueing network model to analyze the impact of parallelization of care on patient cycle time," *Health Care Manage. Sci.*, vol. 11, no. 3, pp. 248–261, 2008.

[35] C. Pehlivan, V. Augusto, X. Xie, and C. Crenn-Hebert, "Multi-period capacity planning for maternity facilities in a perinatal network: A queuing and optimization approach," in *Proc. IEEE Int. Conf. Autom. Sci. Eng.*, Aug. 2012, pp. 137–142.

[36] J. Wang, S. Quan, J. Li, and A. M. Hollis, "Modeling and analysis of work flow and staffing level in a computed tomography division of University of Wisconsin Medical Foundation," *Health Care Manage. Sci.*, vol. 15, no. 2, pp. 108–120, 2012.

[37] J. Wang, X. Zhong, J. Li, and P. K. Howard, "Modeling and analysis of care delivery services within patient rooms: A system-theoretic approach," *IEEE Trans. Autom. Sci. Eng.*, vol. 11, no. 2, pp. 379–393, Apr. 2014.

[38] X. Zhong, J. Song, J. Li, S. M. Ertl, and L. Fiedler, "Design and analysis of gastroenterology (GI) clinic in Digestive Health Center of University of Wisconsin Health," *Flexible Services Manuf. J.*, vol. 28, no. 1, pp. 90–119, 2016.

[39] X. Zhong, J. Li, S. M. Ertl, C. Hassemer, and L. Fiedler, "A system-theoretic approach to modeling and analysis of mammography testing process," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 1, pp. 126–138, Jan. 2016.

[40] B. Zeng, A. Turkcan, J. Lin, and M. Lawley, "Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities," *Ann. Oper. Res.*, vol. 178, no. 1, pp. 121–144, 2010.

[41] W.-Y. Wang and D. Gupta, "Adaptive appointment systems with patient preferences," *Manuf. Service Oper. Manage.*, vol. 13, no. 3, pp. 373–389, 2011.

[42] C. Zacharias and M. Pinedo, "Appointment scheduling with no-shows and overbooking," *Prod. Oper. Manage.*, vol. 23, no. 5, pp. 788–801, 2014.

[43] G. Kaplan, M. H. Lopez, and J. M. McGinnis, Eds., *Transforming Health Care Scheduling and Access: Getting to Now*. Washington, DC, USA: National Academies Press (US), Aug. 2015. [Online]. Available: http://www.ncbi.nlm.nih.gov/books/NBK316132/, doi: 10.17226/20220.

[44] S. A. Shipman and C. A. Sinsky, "Expanding primary care capacity by reducing waste and improving the efficiency of care," *Health Affairs*, vol. 32, no. 11, pp. 1990–1997, 2013.

[45] R. B. Cooper, *Introduction to Queueing Theory*, 2nd ed. New York, NY, USA: North Holland, 1981.

[46] H. Takagi, *Queueing Analysis: Vacation and Priority Systems*, vol. 1. Amsterdam, The Netherlands: North Holland, 1991.

Xiang Zhong received the B.S. degree from the Department of Automation, Tsinghua University, Beijing, China, in 2011, and the M.S. degree in statistics from the University of Wisconsin–Madison, Madison, WI, USA, in 2014, where she is currently pursuing the Ph.D. degree with the Department of Industrial and Systems Engineering.

She has been an Assistant Professor with the Department of Industrial and Systems Engineering, University of Florida, Gainesville, FL, USA, since 2016. Her current research interests include stochastic modeling and data analytics with the application in healthcare, service, and production systems.

Jingshan Li (SM'97–M'00–SM'06) received the B.S. degree from the Department of Automation, Tsinghua University, Beijing, China, in 1989, the M.S. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, in 1992, and the Ph.D. degree in electrical engineering-systems from the University of Michigan, Ann Arbor, MI, USA, in 2000.

He was a Staff Research Engineer with the Manufacturing Systems Research Laboratory, General Motors Research and Development Center, Warren, MI, USA, from 2000 to 2006, and was with the Department of Electrical and Computer Engineering and the Center for Manufacturing, University of Kentucky, Lexington, KY, USA, from 2006 to 2010. He is currently a Professor with the Department of Industrial and Systems Engineering, University of Wisconsin–Madison, Madison, WI, USA. His current research interests include modeling, analysis, and control of manufacturing and healthcare systems.

Prof. Li received the 2010 NSF Career Award, the 2009 IIE Transactions Best Application Paper Award, the 2005 IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING Best Paper Award, the 2006 IEEE Early Industry/Government Career Award in Robotics and Automation, and multiple awards in international conferences. He is a Senior Editor of the IEEE ROBOTICS AND AUTOMATION LETTERS, the Department Editor of *IIE Transactions*, and an Associate Editor of the IEEE TRANSACTIONS ON AUTOMATION SCIENCE AND ENGINEERING, the *International Journal of Production Research*, *Flexible Service and Manufacturing*, and the *International Journal of Automation Technology*.

Philip A. Bain received the M.D. degree from the University of Wisconsin Medical School, Madison, WI, USA, in 1895.

He completed his residency in Internal Medicine with the Medical College of Wisconsin Affiliated Hospitals, Milwaukee, WI, USA, in 1988. He is currently the Chief of Internal Medicine with Dean Health System, Inc., Madison. He is a Practicing General Internist and a Fellow of the American College of Physicians. His current research interests include practice efficiency, work flow analysis, and population health.

Albert J. Musa received the bachelor's degree from Augustana College, Rock Island, IL, USA, in 1979, completed medical school at Rush Medical College, Chicago, IL, USA, in 1983, completed residencies in both family medicine and internal medicine at the University of Wisconsin–Madison, Madison, WI, USA, in 1986 and 1988, respectively, and is board certified in both. He completed Six Sigma Black Belt Certification at UW Executive Education in 2011. He received the Master of Health Care Delivery Science degree from Dartmouth College, Hanover, NH, USA, in 2016.

He is an MD and has been a practicing general internist since 1987 with Dean Health System. He has held a number of part time administrative positions including Chief of Staff at St. Marys Hospital-Madison, voluntary teaching appointment at UW-Madison School of Medicine. He is actively involved in the Wisconsin Research and Education Network (a practice-based research network) and i-Practise, (Improving Primary Care through Industrial and Systems Engineering), both representing academic and community collaborative partnerships. He served as the Medical Director of the Primary Care East Region of Dean Health System (DHS), Inc., in Madison, Wisconsin from 2007 to 2014. He is currently a Board Director for the Dean Health System Board of the SSM WI Integrated Health System Board.