Decision Support Systems 53 (2012) 554–564

Contents lists available at SciVerse ScienceDirect

Decision Support Systems

journal homepage: www.elsevier.com/locate/dss

A mean–variance model to optimize the fixed versus open appointment percentages in open access scheduling systems

Xiuli Qu [a,*], Ronald L. Rardin [b], Julie Ann S. Williams [c]

[a] North Carolina Agricultural and Technical State University, Department of Industrial and Systems Engineering, 1601 E. Market Street, Greensboro, NC 27411, United States
[b] University of Arkansas, Department of Industrial Engineering, 4169 Bell Engineering Center, Springdale, AR 72767, United States
[c] University of West Florida, Department of Management and MIS, 11000 University Parkway, Bldg 76/128, Pensacola, FL 32514-5752, United States

ARTICLE INFO

Article history:
Received 3 March 2011
Received in revised form 26 December 2011
Accepted 15 April 2012
Available online 24 April 2012

Keywords:
Appointment scheduling
Health care policy
Service operations
Open access scheduling
Mean–variance model

ABSTRACT

Although healthcare quality may improve with short-notice scheduling and subsequently higher patient show-up rates, the variability in patient flow may negatively impact the service design. This study demonstrates how to select the percentage for short-notice or open appointments in an open access scheduling system subject to two quality performance metrics. Specifically, we develop a mean–variance model and an efficient solution procedure to help clinic administrators determine the open appointment percentage subject to increasing the average number of patients seen while also reducing the variability. Our numerical results indicate that for cases with high patient demand and high patient no-show rates for fixed appointments, one or more Pareto optimal percentages of open appointments significantly decrease the variability in the number of patients seen with only a negligible decrease in the expected number of patients seen. While our method provides a useful tool for clinic administrators, it also presents a modeling foundation for open access scheduling with quality management objectives to smooth patient flow and improve capacity utilization.

1. Introduction

Health expenditures have grown rapidly in all developed countries. Faced with the dual pressures of reducing costs and improving quality of care, hospitals have not only improved their operations but they have also shifted care from inpatient facilities to outpatient clinics since the 1980s [21,42]. Subsequently, outpatient primary care clinics have been forced to improve their just-in-time accessibility and use of limited capacity. A significant problem for primary care clinics is the traditional appointment scheduling system, which schedules patients' routine check-ups months in advance. When a healthcare delivery system is viewed as a production system in which the deliverables are patients receiving healthcare services, production concepts related to scheduling and capacity management can be adapted to improve allocation of provider capacity. Healthcare delivery research is making important contributions to optimize elective surgery schedules under uncertain demand for emergency surgeries [8,20]. Unlike surgical appointment scheduling, primary care outpatient scheduling experiences significantly high patient no-shows while being constrained by a lower degree of flexibility to increase capacity into emergency hours with emergency staff.

While over fifty years of appointment scheduling research has focused on traditional appointment scheduling systems and its parameters [9,18,19,23–25,33,47] (see the review by Cayirli and Veral [5] and the review by Gupta and Denton [14]), our paper makes important contributions to determine the parameters for the new short-notice scheduling systems. The key principle of short-notice scheduling is to see patients when they want to be seen, similar to the idea of just-in-time in the manufacturing industry. Throughout the paper, short-notice appointments are called *open appointments*, while appointments scheduled in advance are called *fixed appointments*. While scheduling a routine check-up months in advance accommodates patients who need longer lead times to accommodate transportation, work, family, or pre-test dietary intake arrangements, there are also several key costs that include a scarcity of appointment slots for short-notice acute care appointments and historically high patient no-show rates [2,11,22]. Therefore, the short-notice scheduling, called *open access scheduling*, *advanced access scheduling* or *same-day scheduling*, redesigns outpatient appointment scheduling systems to provide short-notice appointments for both routine check-ups and acute illnesses. Reports of successful implementation of open access scheduling in primary care clinics demonstrate its advantages: reduced patient no-show rates and cost per service as well as improved continuity of care, patient satisfaction, operational efficiency and productivity of healthcare providers, including physicians, nurse practitioners, or physician's assistants [1,4,15,17,30,35,36,40,44]. However, some failure stories to implement open access

* Corresponding author. Tel.: +1 336 334 7780x546; fax: +1 336 334 7729.
E-mail addresses: xqu@ncat.edu (X. Qu), rrardin@uark.edu (R.L. Rardin), JAWilliams@uwf.edu (J.A.S. Williams).

0167-9236/$ – see front matter © 2012 Elsevier B.V. All rights reserved.
doi:10.1016/j.dss.2012.04.003

scheduling point out the difficulties in determining open access scheduling parameters when patient demand is high [29,31].

The critical parameters for open access scheduling systems are reported to be determined by expert's experiences rather than analytical methods. For example, the percentage of open appointments may range from 30% to 80% depending on the experiences of the expert managing the open access clinics [15,17,30]. Qu et al. [38] presented an analytical method to determine the percentage of open appointments that maximizes the expected number of patients consulted. Their numerical results demonstrate that the optimal percentage of open appointments to maximize the expected number of patients consulted mainly depends on the ratio of the average demand for open appointments to provider capacity and the ratio of the show-up rates for fixed versus open appointments. However, the variability in the number of patients consulted is not considered. In an outpatient clinic, healthcare providers have to see all patients who show up for their appointments. High variability in the number of patients consulted per session may result in provider overtime in some sessions, while in other sessions it may result in long provider idle time and low capacity utilization which increases clinic operation costs and adversely affects the experiences of the patients and the staff [3,26,34,48]. Our paper extends the aforementioned analytical model to determine the percentage of open appointments in a critical way. We explicitly consider variability. We seek to not only maximize the average number of patients consulted but also minimize the variance in the number of patients consulted.

The structure of the paper is organized as follows. In the next section, we propose the mean–variance model to determine the percentage of open appointments for a healthcare provider, followed by a procedure to search all Pareto optimal solutions to the model in Section 3. Using the procedure, numerical cases are solved in Section 4 to examine the performance improvement using Pareto optimal percentages of open appointments, and the impacts of patient no-show rates and patient demand on the Pareto optimal solutions. The significance of our study is discussed in Section 5. Finally, potential future research is discussed and conclusions are drawn in Section 6.

2. Mean–variance model

A primary care provider typically sees patients in half-day sessions. Since the average number of patients consulted in each session affects the revenues and costs of a clinic as well as patient access, it is one of the key performance measures that concerns healthcare administrators and providers. Another performance measure concerning both healthcare administrators and providers is the variation in the number of patients consulted by each provider in each clinic session. The variation in the number of patients consulted results in clinic overtime and long patient waiting time in some sessions, but long provider idle time in other sessions. In a study [19], LaGanga and Lawrence discussed the challenges of managing varying patient attendance and the tradeoffs of overbooking, provider productivity, patient access, patient waiting time, and provider overtime. Their simulation results demonstrate that the average patient waiting time and the average clinic overtime increase as the variation in patient attendance increases. On the other hand, given the expected number of patients seen in a session, the average provider productivity decreases as the variation in the number of patients consulted in each clinic session increases [19]. The decrease of the average provider productivity is equivalent to the increase of the average provider idle time. The variation in the number of patients consulted in a session is caused by patient no-shows and the variation of patient demand for appointments. However, an appropriate percentage of open appointments could reduce the overall variability in an open access scheduling system. Therefore, the percentage of open appointments is better determined by considering both the expectation and the variance of the number of patients consulted.

2.1. Model assumptions

Recent research suggests that continuity of care can improve patient satisfaction, especially early in the patient relationship and for patients with worse health [39,41]. Some clinics group two to four providers as a provider team in an effort to balance continuity of care with scheduling flexibility. We assume a healthcare provider's (or provider team's) schedule is independent of the schedules of other providers (or provider teams). Thus, we model appointment scheduling for one provider (or provider team).

We assume that the patients independently request appointments and independently choose appointments in other sessions when the desired session is full. Many discrete models for customer choice behavior make a similar assumption about independent customer choice [28,45]. Thus, patient demand in a session is independent of patient demands in other sessions. We assume that for a given provider (or provider team), the joint distribution of demands for fixed and open appointments is known. Demand correlation between open and fixed appointments within the same session may be positive, negative, or independent. In the numerical study section, we examine the sensitivity of Pareto optimal percentages of open appointments to independent, positively correlated, and negatively correlated demand distributions for fixed and open appointments.

Since clinics have fixed hours and a limited number of scheduled providers and exam rooms, the maximum number of patients that could be seen in a session without overtime is known. Then the total number of appointments to be scheduled in a session could be determined based on the maximum number of patients seen in a session, the average no-show rate and the acceptable possibility of clinic overtime [16,43,47]. In addition, we also assume that patient no-shows are independent of each other, and the no-show rates of fixed and open appointments are known. Meanwhile, it is assumed that no-show rates increase with the increase in the interval from the date an appointment is scheduled to the appointment date [2,11,22].

2.2. Formulation

Since the total number of appointments that can be scheduled with a healthcare provider in a session, denoted by N, is known, determining the optimal percentage of open appointments with a provider in a session is equivalent to determining the optimal number of fixed appointments that can be scheduled with the provider in the session (n_1). Thus, the mean–variance model to Pareto optimize the expectation and the variance (equivalently standard deviation) of the number of patients consulted can be formulated

$$
\begin{aligned}
(\mathrm{P}_0)\quad & \text{maximize } E_{n_1}(\underset{\sim}{M}) \\
& \text{minimize } \sqrt{V_{n_1}(\underset{\sim}{M})} \\
& \text{subject to } n_1 \le N \\
& \qquad n_1 \text{ is integer}
\end{aligned}
$$

where $\underset{\sim}{M}$ denotes the random number of patients consulted by a provider in a session, and $E_{n_1}(\underset{\sim}{M})$ and $V_{n_1}(\underset{\sim}{M})$ denote its expectation and variance, respectively, given that at most n_1 fixed appointments can be scheduled. The mean–variance model (P_0) determines Pareto optimal allocations of open versus fixed appointments that maximize the expected number of patients consulted while minimizing the standard deviation of the number of patients consulted. Pareto optimal solutions to the model (P_0) provide more decision options for clinic administrators when they have to make a trade-off between the average number of patients seen and the variability in the number of patients seen.

3. Solution method

3.1. Expectation and variance of the number of patients consulted

To search a Pareto optimal solution to Problem (P_0), we first derive the formulas to determine the expectation and the variance of the number of patients consulted, $E_{n_1}(\underset{\sim}{M})$ and $V_{n_1}(\underset{\sim}{M})$. When a limited number of fixed appointments can be scheduled with a provider in a session, the number of patients consulted ($\underset{\sim}{M}$) is a function of the total number of appointments available (N), the limit of fixed appointments to be scheduled (n_1), the no-show rates of fixed and open appointments (denoted by γ_1 and γ_2), and the demand distribution for fixed and open appointments (denoted by $\underset{\sim}{D}_1$ and $\underset{\sim}{D}_2$). The joint probability mass function of demands $\underset{\sim}{D}_1$ and $\underset{\sim}{D}_2$ is $p(d_1,d_2) = P\left(\underset{\sim}{D}_1 = d_1, \underset{\sim}{D}_2 = d_2\right)$ for $d_1 = 0, 1, 2, \cdots$ and $d_2 = 0, 1, 2, \cdots$.

Let $\underset{\sim}{M}_1^{(n_1)}$ and $\underset{\sim}{M}_2^{(n_1)}$ denote the random numbers of fixed and open appointments scheduled, respectively, with a provider in a clinic session. We know $\underset{\sim}{M}_1^{(n_1)} = \min(n_1, \underset{\sim}{D}_1)$ and $\underset{\sim}{M}_2^{(n_1)} = \min(N-\underset{\sim}{M}_1^{(n_1)}, \underset{\sim}{D}_2)$. Define independent random variables

$$\underset{\sim}{X}_{1i}^{(n_1)} = \begin{cases} 1, \text{if the patient with the i}^{\text{th}}\text{fixed appointment in a session shows up} \\ 0, \text{otherwise} \end{cases}, \tag{1}$$

and

$$\underset{\sim}{X}_{2i}^{(n_1)} = \begin{cases} 1, \text{if the patient with the i}^{\text{th}}\text{open appointment in a session shows up} \\ 0, \text{otherwise} \end{cases}, \tag{2}$$

where $P(\underset{\sim}{X}_{1i}^{(n_1)} = 1) = 1-\gamma_1$ and $P(\underset{\sim}{X}_{2i}^{(n_1)} = 1) = 1-\gamma_2$. Thus

$$\underset{\sim}{M} = \sum_{i=1}^{\underset{\sim}{M}_1^{(n_1)}} \underset{\sim}{X}_{1i}^{(n_1)} + \sum_{i=1}^{\underset{\sim}{M}_2^{(n_1)}} \underset{\sim}{X}_{2i}^{(n_1)} \tag{3}$$

Therefore, $E_{n_1}(\underset{\sim}{M})$ and $V_{n_1}(\underset{\sim}{M})$ are also functions of N, n_1, γ_1, γ_2, and $p(d_1,d_2)$. Given N, γ_1, γ_2, and $p(d_1, d_2)$, Propositions 1 and 2 present recurrence relations to efficiently calculate $E_{n_1}(\underset{\sim}{M})$ and $V_{n_1}(\underset{\sim}{M})$ for all $n_1 \in \{0, 1, \ldots, N\}$, respectively.

Proposition 1. *Given N, n_1, γ_1, γ_2, and $p(d_1,d_2)$, the expected number of patients consulted by a provider in a clinic session can be determined by a recurrence relation*

$$\begin{aligned} E_{n_1}(\underset{\sim}{M}) &= E_{n_1-1}(\underset{\sim}{M}) - (\gamma_1-\gamma_2)[1-F_1(n_1-1)] \\ &\quad + (1-\gamma_2)[F_2(N-n_1) - F(n_1-1, N-n_1)], for\ 0<n_1\leq N, \end{aligned} \tag{4}$$

with

$$E_0(\underset{\sim}{M}) = (1-\gamma_2)\left[N - \sum_{d_2=0}^{N}(N-d_2)p_2(d_2)\right], \tag{5}$$

where $F_1(\bullet)$, $F_2(\bullet)$, and $F(\bullet)$ are the marginal and joint cumulative probability distribution functions of $\underset{\sim}{D}_1$ and $\underset{\sim}{D}_2$, respectively, and $p_2(\bullet)$ is the probability mass function of $\underset{\sim}{D}_2$.

Proof. See Appendix A.

Proposition 2. *Given N, n_1, γ_1, γ_2, and $p(d_1,d_2)$, the variance of the number of patients consulted by a provider in a session can be determined by a recurrence relation*

$$\begin{aligned} V_{n_1}(\underset{\sim}{M}) &= V_{n_1-1}(\underset{\sim}{M}) + \gamma_1(1-\gamma_1)[1-F_1(n_1-1)] \\ &\quad + (1-\gamma_1)^2[1-F_1(n_1-1)][C_1(n_1-1)+C_1(n_1)] \\ &\quad - \gamma_2(1-\gamma_2)G(n_1-1, N-n_1) - (1-\gamma_2)^2 G(n_1-1, N-n_1)[C_2(n_1-1) \\ &\quad + C_2(n_1)] + 2(1-\gamma_1)(1-\gamma_2)\{G(n_1-1, N-n_1)[N-n_1-C_1(n_1-1)] \\ &\quad - [1-F_1(n_1-1)][N-n_1-C_2(n_1)] + C_3(n_1-1, N-n_1)\},\ \ for\ 0<n_1\leq N, \end{aligned} \tag{6}$$

with

$$\begin{aligned} V_0(\underset{\sim}{M}) &= \gamma_2(1-\gamma_2)[N-C_2(0)] \\ &\quad + (1-\gamma_2)^2\left\{\sum_{d_2=0}^{N}(N-d_2)^2 p_2(d_2) - [C_2(0)]^2\right\}. \end{aligned} \tag{7}$$

Here $G(n_1-1,\ N-n_1) = 1 - F_1(n_1-1) - F_2(N-n_1) + F(n_1-1,\ N-n_1)$, $C_1(n_1) = n_1 - E\left[\underset{\sim}{M}_1^{(n_1)}\right]$, $C_2(n_1) = N - n_1 - E\left[\underset{\sim}{M}_2^{(n_1)}\right]$ and $C_3(n_1-1, N-n_1) = \sum_{d_2=0}^{N-n_1} d_2 p_2(d_2) - \sum_{d_1=0}^{n_1-1}\sum_{d_2=0}^{N-n_1} d_2 p(d_1,d_2)$, where $E\left[\underset{\sim}{M}_1^{(n_1)}\right]$ and $E\left[\underset{\sim}{M}_2^{(n_1)}\right]$, the expectations of $\underset{\sim}{M}_1^{(n_1)}$ and $\underset{\sim}{M}_2^{(n_1)}$, can be calculated, respectively, by

$$E\left[\underset{\sim}{M}_1^{(n_1)}\right] = E\left[\underset{\sim}{M}_1^{(n_1-1)}\right] + [1-F_1(n_1-1)], \tag{8}$$

and

$$E\left[\underset{\sim}{M}_2^{(n_1)}\right] = E\left[\underset{\sim}{M}_2^{(n_1-1)}\right] - G(n_1-1, N-n_1). \tag{9}$$

Proof. See Appendix A.

3.2. Procedure to search Pareto optimal solutions

According to the recurrence relations in Propositions 1 and 2, a procedure is developed to find all Pareto optimal solutions to Problem (P_0). Fig. 1 illustrates the steps in this procedure. In the procedure, the first step is to calculate the marginal probabilities and the joint cumulative probabilities of the demands for fixed and open appointments. In the second step, the expectations and the variances of the number of patients consulted, $E_{n_1}(\underset{\sim}{M})$ and $V_{n_1}(\underset{\sim}{M})$, for all possible n_1, are calculated using Eqs. (4)–(7). The last step is to find all Pareto optimal solutions by sorting and comparing all pairs of $E_{n_1}(\underset{\sim}{M})$ and $V_{n_1}(\underset{\sim}{M})$. The detailed procedure is provided in Appendix B.

In this procedure, calculating the probabilities needed takes time $O(N^2)$, where N is the total number of appointments to be scheduled with a provider (or a provider team) in a clinic session. After that, calculating $E_{n_1}(\underset{\sim}{M})$ and $V_{n_1}(\underset{\sim}{M})$ for all possible n_1 takes time $O(N^2)$. Lastly, sorting all pairs of $E_{n_1}(\underset{\sim}{M})$ and $V_{n_1}(\underset{\sim}{M})$ takes time $O(N\log N)$, and identifying all Pareto optimal solutions takes time $O(N)$. The procedure can quickly find all Pareto optimal limits for the number of fixed appointments to be scheduled in a primary care clinic of typical size. For example, Procedure 1 is coded using MATLAB 7.0.4 to search all Pareto optimal solutions. The computation time to find all Pareto optimal solutions for any representative numerical case in the next section is less than 1 second on a DELL Pentium IV 2.8 G personal computer.

4. Numerical study

Using 540 representative numerical cases, we investigate the two questions of interest to the healthcare administrator: (1) Compared to the percentage of open appointments that maximizes the expected number of patients consulted, does the mean–variance model lead to better decisions on the allocation of open versus fixed appointments? (2) How sensitive are the Pareto optimal allocations of open versus fixed appointments to the patient demands and patient no-show rates? These questions focus on performance improvement using Pareto optimal allocations of open versus fixed appointments, and the impacts of patient characteristics.

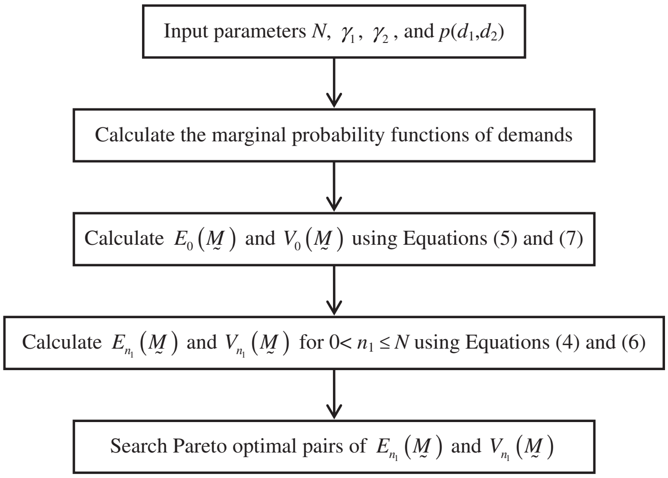

Fig. 1. Flowchart of the procedure to find all Pareto optimal solutions to Problem (P_0).

4.1. Characteristics of numerical cases

The characteristics of the numerical cases were based on real-life examples from clinics visited. In an outpatient clinic, the total number of appointments to be scheduled represents the daily capacity of a provider. The clinics visited had 4-hour sessions in which 10–20 appointments were scheduled with a provider in a session. To examine the impact of the total number of appointments available, three levels are tested in the numerical cases: 12, 16 and 24.

Due to the national shortage of primary care providers, there are many primary care clinics where patient demand exceeds provider capacity [12,13,27]. To manage excess demand, clinic administrators and physicians control the patient panel size of each physician by limiting the number of new patients. Meanwhile, many clinics adopted open access scheduling to increase provider utilization by reducing patient no-show rates and by accepting more new patients [15,36,46]. Therefore, three levels of total average patient demand relative to provider capacity are considered in the numerical cases: $E(\underset{\sim}{D}_1+\underset{\sim}{D}_2)=0.8N$, $E(\underset{\sim}{D}_1+\underset{\sim}{D}_2)=N$ and $E(\underset{\sim}{D}_1+\underset{\sim}{D}_2)=1.2N$. In addition, to examine the impact of equal and unequal demands for fixed and open appointments, three demand allocations, $E(\underset{\sim}{D}_1){:}E(\underset{\sim}{D}_2)=1{:}4$, 1:1 and 4:1, are assumed in the numerical cases. Since clinics experience seasonal surges in demand such as flu season, testing three demand allocations will provide insights into whether the capacity allocation should be adjusted for seasonality.

Since the demand for fixed or open appointments is the number of requests for appointments occurring in a given time period, a Poisson distribution is a routine choice for the demand. When all fixed appointment slots in a session are full, some patients may call back later for an open appointment in their desired session. This results in the positive correlation between the demands for fixed and open appointments in a session. On the other hand, if the patient panel size is assumed to be stable, the decrease in the demand for fixed appointments leads to the increase in the demand for open appointments. That means that the demand for open appointments in a session may be negatively correlated with that for fixed appointments in the session. Therefore, in the numerical cases, the demands for fixed and open appointments are assumed to have independent Poisson distributions or a bivariate Poisson distribution with correlation coefficient of −0.4, −0.2, 0.2 or 0.4.

Since fixed appointments can be scheduled weeks in advance, their no-show rates can reach as high as 50–55% according to the literature [11,22]. On the other hand, the no-show rate of open appointments reported in the literature is 3–16% [4]. In the study of Qu et al. [38], four combinations of the no-show rates of fixed and open appointments are investigated: (0.1556, 0.05), (0.2, 0.1), (0.3667, 0.05), and (0.4, 0.1). For comparison, the same combinations of the no-show rates are considered in the numerical cases here. Table 1 summarizes the levels of provider capacity (N), the no-show rates (γ_1,γ_2), and the demand distribution for fixed and open appointments. Combining these levels generates the 540 numerical cases.

4.2. Performance of optimal allocations determined by the mean–variance model

As mentioned earlier, Qu et al. [38] present a method to determine the percentage of open appointments that maximizes the expected number of patients consulted. Let π^U denote the percentage maximizing the expected number of patients consulted, which is an optimal solution to the model presented in the study of Qu et al. [38] and one of the Pareto optimal solutions to the mean–variance model (P_0). For any case, π^U leads to a higher expectation and a higher standard deviation of the number of patients consulted than other Pareto optimal percentages of open appointments. However, in many numerical cases, some Pareto optimal percentage significantly decreases the standard deviation while only slightly decreasing the expectation of the number of patients consulted. In such cases, the mean–variance model (P_0) provides better decision-making options for the allocation of open versus fixed appointments. Fig. 2 illustrates the performance comparison between π^U and the Pareto optimal percentages of open appointments in terms of the relative decreases in the standard deviation and the expectation of the number of patients consulted. Fig. 2(a) shows that in more than 80% of 540 numerical cases, the percentage of the decrease in the standard deviation of the number of patients consulted is greater than the percentage of the decrease in the expected number of patients consulted. In these cases, the average relative decrease in the standard deviation is 5.82%, and the

Table 1
Levels of clinic and patient characteristics in numerical cases.

Clinic characteristics		Levels	
Total number of appointments to be scheduled (N), i.e. provider capacity		12, 16, and 24	
No-show rates of fixed and open appointments (γ_1, γ_2)		(0.1556, 0.05), (0.2, 0.1), (0.3667, 0.05), and (0.4, 0.1)	
Patient demand	Distribution of $\underset{\sim}{D}_1$ and $\underset{\sim}{D}_2$	Independent Poisson distributions and Bivariate Poisson distributions with correlation coefficient (ρ) of −0.4, −0.2, 0.2 and 0.4	
	Average demands for fixed and open appointments ($E(\underset{\sim}{D}_1)$, $E(\underset{\sim}{D}_2)$)	$E(\underset{\sim}{D}_1+\underset{\sim}{D}_2)=0.8N$	(0.16N, 0.64N), (0.4N, 0.4N), (0.64N, 0.16N)
		$E(\underset{\sim}{D}_1+\underset{\sim}{D}_2)=N$	(0.2N, 0.8N), (0.5N, 0.5N), (0.8N, 0.2N)
		$E(\underset{\sim}{D}_1+\underset{\sim}{D}_2)=1.2N$	(0.24N, 0.96N), (0.6N, 0.6N), (0.96N, 0.24N)

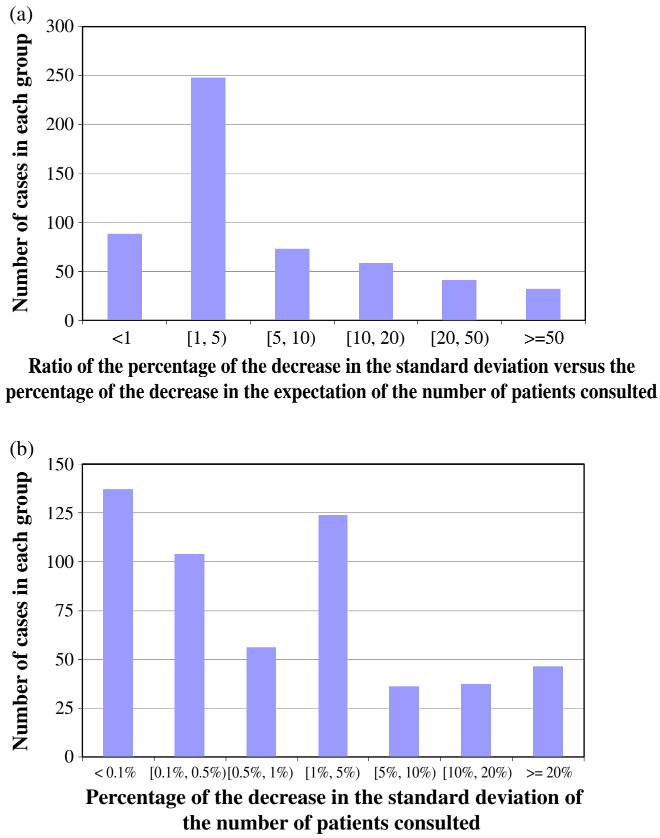

Fig. 2. Performance comparison between Pareto optimal percentages of open appointments and the percentage that only maximizes the expected number of patients consulted.

magnitude of average decrease in the standard deviation is 0.16. The results of LaGanga and Lawrence [19] implies that the decrease of 0.16 in the standard deviation, on average, results in a decrease of about 4 min in the average patient waiting time and a decrease of about 7 min in the average clinic overtime. Fig. 2(b) demonstrates that the standard deviation of the number of patient consulted decreases by less than 0.1% in about 25% of the 540 numerical cases, by 0.1%–1% in about 30% of all cases, by 1%–5% in about 23% of all cases, and by more than 5% in about 22% of all cases. Furthermore, Fig. 2(a) also reveals that in about 25% of the 540 numerical cases, the percentage of the decrease in the standard deviation is more than 10 times greater than the percentage of the decrease in the expected number of patients consulted.

Table 2 illustrates 22 cases in which a Pareto optimal percentage decreases the standard deviation by at least 3% while decreasing the expectation by at most 0.6%. In one of the cases highlighted in Table 2, a Pareto optimal percentage decreases the standard deviation by 17.22% while decreasing the expectation only by 0.59%. In all 22 cases, the total average patient demand is higher than the provider capacity, and the no-show rate of fixed appointments is at the high level (i.e. 36.67% or 40%). This implies that for a case with high patient demand and a high no-show rate of fixed appointments, one or more Pareto optimal percentages of open appointments may significantly decrease the variation in the number of patients consulted while only slightly decreasing the expected number of patients consulted. Therefore, for such clinics the mean–variance model could provide better decision options for the allocation of open versus fixed appointments.

4.3. *Sensitivity of Pareto optimal solutions to patient demand and no-show rates*

Before discussing the details of the sensitivity analysis, we begin with a discussion of two of the typical mean–variance curves from the 540 numerical cases. Fig. 3(a) illustrates the mean–variance curve and Pareto optimal solutions for a case with provider capacity $N=12$, no-show rates $\gamma_1=0.3667$ and $\gamma_2=0.05$, and positively correlated demands for fixed and open appointments. Each point on the curve depicts the expectation and the standard deviation of the number of patients consulted for one feasible solution of the case. For this case, the feasible solutions are $n_1=0, 1, \ldots, 12$, where n_1 is the number of fixed appointments allowed to be scheduled. Since some feasible solutions, such as $n_1=10$, 11 and 12, have very close expectations and standard deviations, the points (markers) corresponding to these feasible solutions overlap in Fig. 3(a). Solid points on the mean–variance curve represent Pareto optimal solutions, which are not dominated by any other feasible solution. Similarly, Fig. 3(b) illustrates the mean–variance curve and Pareto optimal solutions for another case with provider capacity $N=12$, no-show rates $\gamma_1=0.3667$ and $\gamma_2=0.05$, and positively correlated demands for fixed and open appointments. The difference between the two cases is that the case in Fig. 3(b) has a total average patient demand 20% higher than provider capacity while the case in Fig. 3(a) has a total average patient demand 20% lower than provider capacity. Next, we discuss the impact of patient demand and no-show rates on the mean–variance curves and the Pareto optimal solutions.

Fig. 4 demonstrates that the mean–variance curve shape and the efficient frontier shape for each case do not depend solely on provider capacity. Instead, the mean–variance curve shape and the efficient frontier shape depend on the relationship between average patient demand and provider capacity. Since the average patient demands in the numerical cases are determined based on the relationship between average patient demand and provider capacity, we compare the mean–variance curves and Pareto optimal solutions for 72 cases with the same provider capacity $N=16$, which are illustrated in Fig. 5. For the cases in Fig. 5(a), (c) and (e), the total average demand for fixed and open appointments is 20% lower than provider capacity, while for the cases in Fig. 5(b), (d) and (f), the total average demand is 20% higher than provider capacity. Meanwhile, the ratio of the average demands for fixed versus open appointments is 1:1 for the cases in Fig. 5(a) and (b), is 1:4 for the cases in Fig. 5(c) and (d), is 4:1 for the cases in Fig. 5(e) and (f).

Fig. 5(a), (c) and (e) demonstrates that for the cases with patient demand lower than provider capacity, the average patient demands and the demand correlation for fixed and open appointments dominates the shapes of mean–variance curves and efficient frontiers. Compared with the average patient demand and the demand correlation, the no-show rates of fixed and open appointments have less significant effect on the shapes of mean–variance curves and efficient frontiers. Fig. 5(e) reveals that for the cases with patient demand lower than provider capacity and the average demand for open appointments lower than that for fixed appointments, the efficient frontier includes all feasible solutions. For a few cases in Fig. 5(a), the efficient frontier also includes all feasible solutions. The reason is that when provider capacity is much higher than patient demand, most requests for appointments are granted, which results in the simultaneous increases in the expectation and the standard deviation of the number of patients consulted. For such cases, the percentage of open appointments has to be determined by making a trade-off between the expectation and the variance of the number of patients consulted. One way to make this trade-off is to choose the percentage of open appointments for a provider (or a provider team) that maximizes the expectation while ensuring that the variance does not exceed the desired value. However, Fig. 5(c) shows that for the cases with patient demand lower than provider capacity and the average

Table 2
Numerical cases in which a Pareto optimal percentage significantly decreases the standard deviation while slightly decreasing the expectation of the number of patients consulted.

Provider capacity (N)	Patient demand			No-show rate		Using π^U		Using a Pareto optimal percentage		Percentage of the decrease	
	$E(\underset{\sim}{D}_1)$	$E(\underset{\sim}{D}_2)$	ρ	γ_1	γ_2	Expected number	Standard deviation	Expected number	Standard deviation	Expected number	Standard deviation
12	7.2	7.2	−0.2	0.3667	0.05	8.946	1.728	8.937	1.666	0.11%	3.57%
				0.4	0.1	8.476	1.768	8.467	1.714	0.11%	3.06%
16	3.84	15.36	0	0.3667	0.05	13.654	1.953	13.604	1.734	0.37%	11.23%
				0.4	0.1	12.936	2.026	12.888	1.838	0.37%	9.31%
			0.2	0.3667	0.05	13.639	1.967	13.614	1.837	0.18%	6.59%
				0.4	0.1	12.921	2.037	12.897	1.925	0.18%	5.50%
			−0.4	0.3667	0.05	13.706	1.786	13.670	1.615	0.26%	9.57%
				0.4	0.1	12.984	1.883	12.951	1.739	0.26%	7.68%
	9.6	9.6	0.4	0.3667	0.05	11.956	2.148	11.940	2.060	0.14%	4.09%
				0.4	0.1	11.327	2.177	11.312	2.099	0.14%	3.58%
			−0.4	0.3667	0.05	12.158	1.939	12.151	1.866	0.06%	3.76%
				0.4	0.1	11.518	1.995	11.511	1.932	0.06%	3.18%
24	5.76	23.04	0	0.3667	0.05	20.862	2.300	20.778	2.028	0.40%	11.85%
				0.4	0.1	19.764	2.406	19.745	2.263	0.10%	5.95%
			0.2	0.3667	0.05	20.849	2.312	20.811	2.158	0.18%	6.66%
				0.4	0.1	19.751	2.416	19.716	2.284	0.18%	5.45%
			−0.2	0.3667	0.05	20.870	2.298	20.747	1.955	0.59%	14.93%
				0.4	0.1	19.771	2.404	19.715	2.162	0.29%	10.05%
			−0.4	0.3667	0.05	20.877	2.295	20.753	1.900	0.59%	17.22%
				0.4	0.1	19.779	2.402	19.701	2.093	0.39%	12.84%
	14.4	14.4	0	0.3667	0.05	18.415	2.424	18.406	2.312	0.05%	4.64%
				0.4	0.1	17.446	2.488	17.437	2.390	0.05%	3.94%

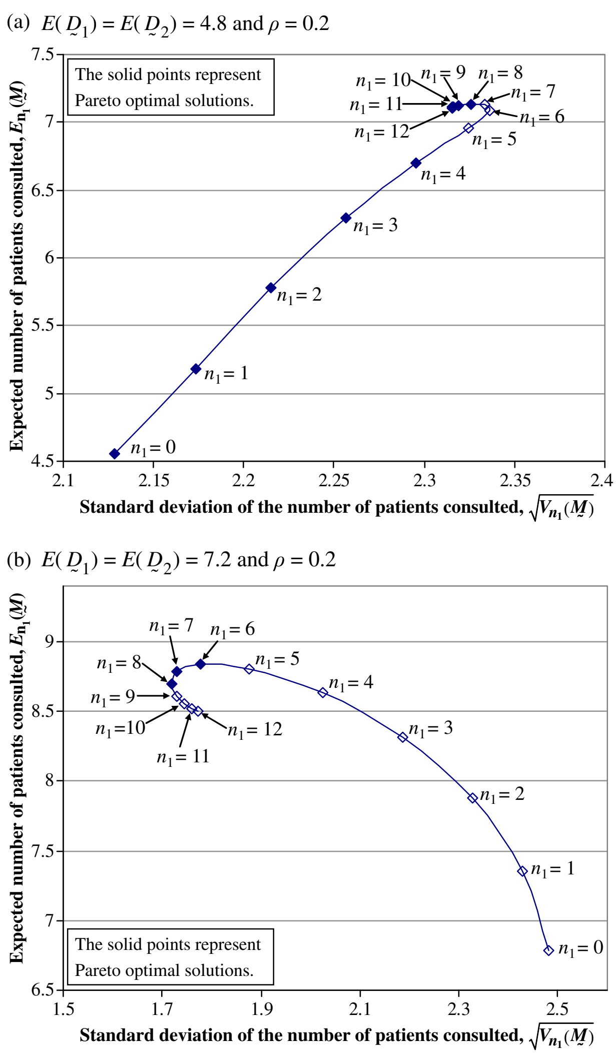

Fig. 3. Mean–variance curves for two cases with $N = 12$, $\gamma_1 = 0.3667$ and $\gamma_2 = 0.05$.

demand for open appointments higher than that for fixed appointments, the standard deviation of the number of patients consulted may decrease with the increase in the expectation.

On the other hand, Fig. 5(b), (d) and (f) demonstrates that for the cases with patient demand higher than provider capacity, the average patient demands for fixed and open appointments dominate the shapes of mean–variance curves and efficient frontiers. The effect of the no-show rates on the shapes of mean–variance curves and efficient frontiers is less significant than the effect of the average demands, but is more significant than the effect of the correlation between the demands for fixed and open appointments. Fig. 5(b) and (d) shows that for the cases with patient demand higher than provider capacity and the average demand for open appointments not less than that for fixed appointments, Pareto optimal percentages

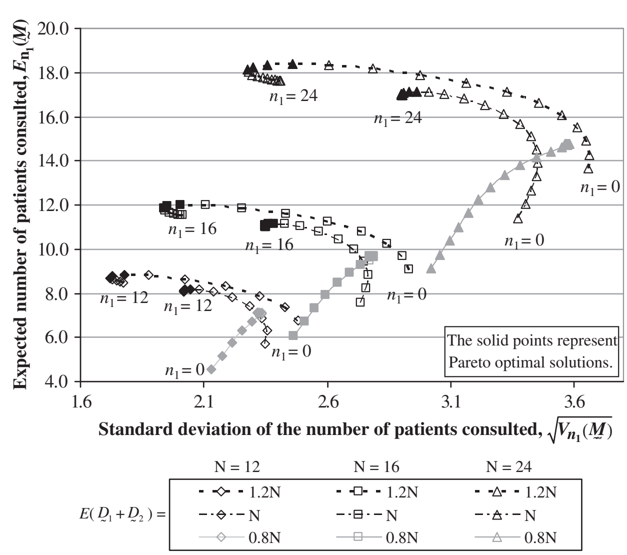

Fig. 4. Mean–variance curves for the cases with $E(\underset{\sim}{D}_1) = E(\underset{\sim}{D}_2)$, $\gamma_1 = 0.3667$, $\gamma_1 = 0.05$, and $\rho = 0.2$ for three provider capacities ($N = 12$, 16, and 24).

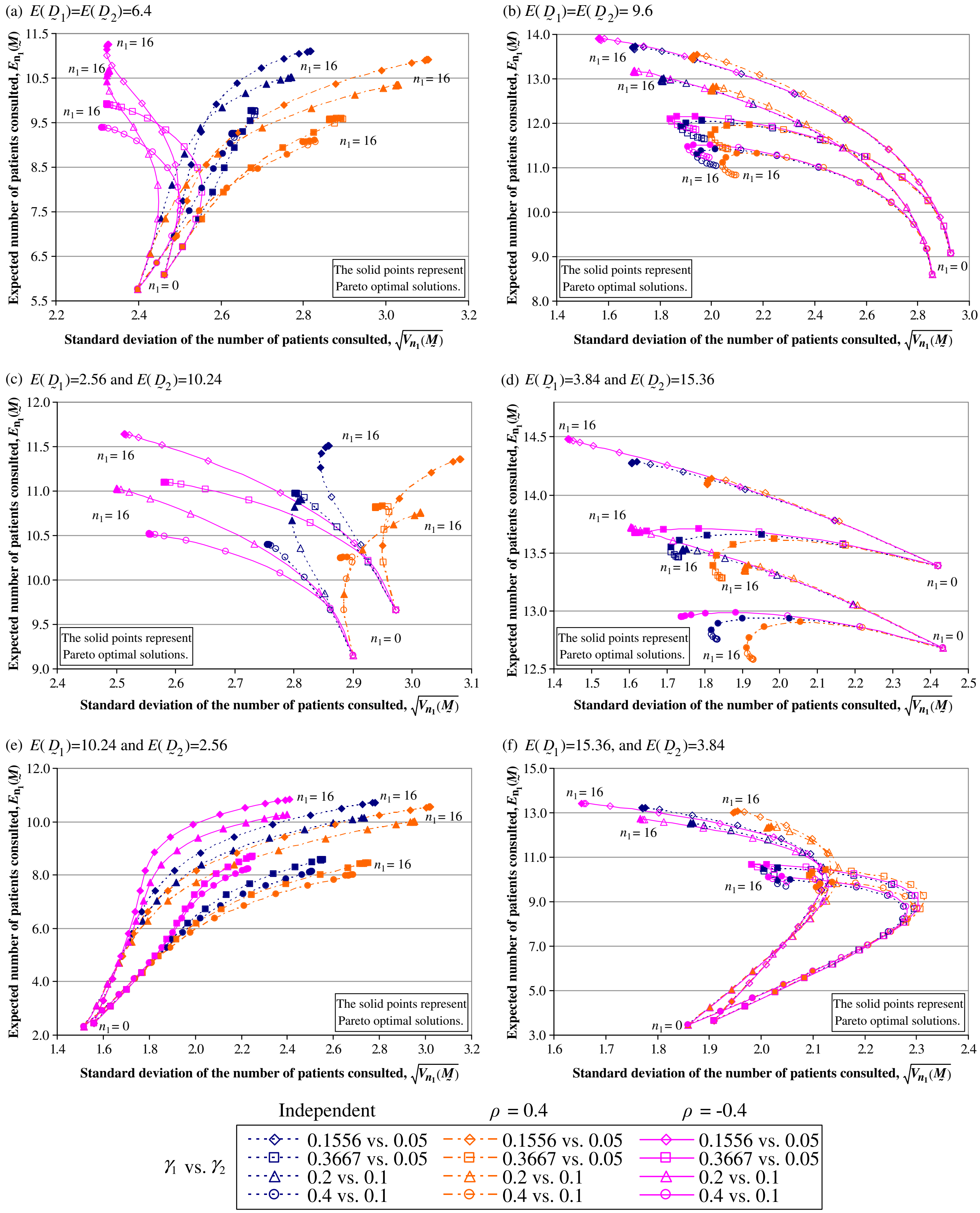

Fig. 5. Mean–variance curves for the cases with $N = 16$.

have very close expectations and standard deviations, or one or more Pareto optimal percentages significantly reduce the standard deviation while only slightly decreasing the expectation. This observation reveals that for such cases the percentage of open appointments that minimizes the standard deviation is a good choice for open access scheduling because it only slightly decreases the expected number of patients consulted. It also supports that for a clinic with high patient demand and high no-show rates for fixed appointments, the mean–variance model could provide better decision options for the allocation of open versus fixed appointments.

5. Discussion

An open access scheduling system requires clinic administrators to make additional new decisions when compared to decision making in traditional appointment scheduling systems. One of these new decisions is the percentage of open appointments, which we have shown in this paper impacts the number of patients seen. In this paper, we propose a mean–variance model and a solution procedure for clinic administrators to determine the percentage of open appointments subject to two objectives: (1) maximizing the average number of patients seen, and (2) minimizing the variability in the number of patients seen. Our numerical results indicate that Pareto optimal solutions to our mean–variance model may decrease the standard deviation by up to 17.22% while decreasing the expectation by only 0.59% when compared to the earlier model in the study of Qu et al. [38]. Meanwhile, the proposed procedure reports all Pareto optimal percentages of open appointments, which provides all of the best options for clinic administrators to determine the percentage of slots reserved for open appointments based on their priorities and concerns. For example, clinic administrators could use our model to reduce the variability in the number of patients seen and to smooth patient flows.

The results of this study also provide insights for clinic administrators to improve the allocation of provider capacity. First, for a clinic with low patient demand for total appointments and either lower demand for open appointments (see Fig. 5(e)), or higher or equal demand for open appointments independent of or positively correlated with demand for fixed appointments (see Fig. 5(a) and (c)), clinic administrators have to determine the percentage of open appointments by making a trade-off between the expected number of patients seen and the variability in the number of patients seen because the expectation and the variance of the number of patients seen increase simultaneously. One way to make this trade-off is to choose the percentage of open appointments for a provider (or a provider team) that maximizes the expectation while ensuring that the variance does not exceed the desired limit. On the other hand, if the clinic has low demand for total appointments as well as higher or equal demand for open appointments negatively correlated with demand for fixed appointments (see Fig. 5(a) and (c)), clinic administrators can optimize the expectation and variance simultaneously. If a clinic has high patient demands for total appointments and open appointments, clinic administrators could choose the percentage of open appointments that minimizes the variability in the number of patients seen, while only negligibly decreasing the average number of patients seen. Our results indicate that clinic administrators should adopt strategies to increase patient demand for open appointments and subsequently improve the performance of an open access scheduling system. The long-term strategy to increase patient demand is to improve the quality of and the access to clinical services [10,44], while there are many short-term strategies such as informing the clinic's patients how to request open appointments, mailing the local community brochures introducing the advantages of open access scheduling, etc. [10].

In addition, the ability to manage the trade-offs between the mean and the variance of the patient flow is critical to the performance of open access scheduling systems. In the open access clinics we visited, clinic administrators always need to make the trade-offs between the average number of patients seen per clinic session and clinic overtime and patient waiting time caused by the uncertainty of patient flow. Meanwhile, the negative impacts of the variation in patient attendance on a scheduling system have been investigated in other studies [3,19,32]. For example, as the variation in patient attendance increases, the average patient waiting time, the average provider idle time, and the average clinic overtime increase [19]. The mean–variance model and the recursive procedure proposed are a useful tool for clinic administrators to make the appropriate trade-offs between the average number of patients seen per clinic session and the negative impact of the variation in patient attendance.

From a research perspective, the results of our research offer evidence that model-based decisions may improve process quality and capacity management, particularly in healthcare [6,37]. Since the quality of care depends on healthcare access and process quality at least as much as clinical quality, model-based decision approaches can help healthcare administrators determine how to allocate their resources [7]. In most capacity management studies, patient waiting cost, provider idle cost and/or overtime cost are considered as objectives. The variation in the number of patient arrivals per clinic session, which directly affects these costs, is considered as one of the objectives in the mean–variance model proposed in this paper. This idea could be applied in other capacity management models to improve decision making in other service sectors. Reducing the variance in healthcare provider utilization may also have positive impacts on medical supply inventories, an interesting supply chain implication for future research.

6. Limitation and conclusions

Our analytical approach required several limiting assumptions. We assumed that the patient demand between sessions is independent. An important research question is how to relax this assumption by considering more than one session for optimization and the impact of correlated patient demands between sessions. While the research presented here assumed that a healthcare provider has a fixed number of slots per session (fixed N), a future direction is to allow limited fluctuation for N. If demand varies widely over time (for example high acute condition appointment demand during flu season or high demand for routine check-ups prior to the start of the academic year), an interesting question is to examine the impact of scheduling if forecasted demand varies and limited capacity modifications are allowed.

Our analytical approach provides a decision-making tool needed by healthcare administrators considering or managing an open access scheduling system. We introduce the first mean–variance model to determine the percentage of open appointments for a provider (or a provider team). Then we prove the recurrence relation for the variance of the number of patients consulted, given the total number of appointments available, the no-show rates, and the demand distributions for fixed and open appointments. According to the recurrence relations for the expectation and the variance of the number of patients consulted, a recursive procedure is proposed to efficiently calculate all expectations and variances, and then a search checks all Pareto optimal solutions to the mean–variance model. When compared to a single objective model that maximizes the patients seen, our mean–variance model can improve clinic operations by reducing the variability in the number of patients seen while maintaining high levels of patient consultations.

In closing, our research extends the first model for open access scheduling by evaluating not only the expected number of patients consulted but also the variance in the number of patients consulted in an effort to help healthcare administrators achieve greater capacity management, patient access, and smoother patient flow through improved patient scheduling. Each of the future research directions will extend the discussion from here to address patient scheduling needs for short lead time and provider continuity.

Acknowledgments

The authors want to thank the administrators, physicians, and staff in the Indiana University Medical Group—Primary Care clinics for providing insights into traditional and open access appointment scheduling in outpatient clinics. The authors are grateful to Professor Bruce W. Schmeiser of Purdue University for the valuable discussions and references. The authors also gratefully acknowledge the support from the Regenstrief Center for Healthcare Engineering.

Appendix A. Proofs

Proof of Proposition 1. The number of patients consulted with a provider in a clinic session is $\underset{\sim}{M}=\sum_{i=1}^{\underset{\sim}{M}_1^{(n_1)}}\underset{\sim}{X}_{1i}^{(n_1)}+\sum_{i=1}^{\underset{\sim}{M}_2^{(n_1)}}\underset{\sim}{X}_{2i}^{(n_1)}$ where $\underset{\sim}{M}_1^{(n_1)}=\min(n_1,\underset{\sim}{D}_1)$, $\underset{\sim}{M}_2^{(n_1)}=\min(N-\underset{\sim}{M}_1^{(n_1)},\underset{\sim}{D}_2)$, and $\underset{\sim}{X}_{1i}^{(n_1)}$ and $\underset{\sim}{X}_{2i}^{(n_1)}$ are independent Bernoulli random variables with $P(\underset{\sim}{X}_{1i}^{(n_1)}=1)=1-\gamma_1$ and $P(\underset{\sim}{X}_{2i}^{(n_1)}=1)=1-\gamma_2$, respectively. Thus, the expected number of patients consulted with the provider in the session is

$$E_{n_1}(\underset{\sim}{M})=(1-\gamma_1)E\left[\underset{\sim}{M}_1^{(n_1)}\right]+(1-\gamma_2)E\left[\underset{\sim}{M}_2^{(n_1)}\right]. \tag{A.1}$$

When the limit of fixed appointments to be scheduled increases from n_1-1 to n_1, the number of fixed appointments scheduled increases by 1 if $\underset{\sim}{D}_1>n_1-1$; otherwise, the number does not change. Thus

$$\begin{aligned}E\left[\underset{\sim}{M}_1^{(n_1)}\right]&=E\left[\underset{\sim}{M}_1^{(n_1-1)}\right]+1\times P\left(\underset{\sim}{D}_1>n_1-1\right)+0\times P\left(\underset{\sim}{D}_1\le n_1-1\right)\\&=E\left[\underset{\sim}{M}_1^{(n_1-1)}\right]+[1-F_1(n_1-1)].\end{aligned} \tag{A.2}$$

On the other hand, when the limit of fixed appointments to be scheduled increases from n_1-1 to n_1, the number of open appointments scheduled decreases by 1 if $\underset{\sim}{D}_1>n_1-1$ and $\underset{\sim}{D}_2\ge N-(n_1-1)$; otherwise, the number does not change. Thus

$$\begin{aligned}E\left[\underset{\sim}{M}_2^{(n_1)}\right]&=E\left[\underset{\sim}{M}_2^{(n_1-1)}\right]+(-1)\times P\left(\underset{\sim}{D}_1>n_1-1\text{and}\underset{\sim}{D}_2\ge N-n_1+1\right)\\&\quad+0\times P\left(\underset{\sim}{D}_1\le n_1-1\text{or}\,\underset{\sim}{D}_2<N-n_1+1\right)\\&=E\left[\underset{\sim}{M}_2^{(n_1-1)}\right]-[1-F_1(n_1-1)-F_2(N-n_1)+F(n_1-1,N-n_1)].\end{aligned} \tag{A.3}$$

By substituting Eqs. (A.2) and (A.3) into Eq. (A.1), Eq. (4) is proven.

Next, we derive the initial condition for calculating $E_{n_1}(\underset{\sim}{M})$. When $n_1=0$, $\underset{\sim}{M}_1^{(0)}=0$ and $\underset{\sim}{M}_2^{(0)}$ equals the minimum of $\underset{\sim}{D}_2$ and N. Thus

$$\begin{aligned}E\left[\underset{\sim}{M}_2^{(0)}\right]&=E\left[\min\left(\underset{\sim}{D}_2,N\right)\right]=\sum_{d_2=0}^{N}d_2p_2(d_2)+\sum_{d_2=N+1}^{\infty}Np_2(d_2)\\&=N-\sum_{d_2=0}^{N}(N-d_2)p_2(d_2).\end{aligned} \tag{A.4}$$

Substituting $E\left[\underset{\sim}{M}_1^{(0)}\right]=0$ and Eq. (A.4) into Eq. (A.1), we obtain Eq. (5).

Proof of Proposition 2. Let $Var\left[\underset{\sim}{M}_1^{(n_1)}\right]$ and $Var\left[\underset{\sim}{M}_2^{(n_1)}\right]$ denote the variances of $\underset{\sim}{M}_1^{(n_1)}$ and $\underset{\sim}{M}_2^{(n_1)}$, respectively, and $Cov\left[\underset{\sim}{M}_1^{(n_1)},\underset{\sim}{M}_2^{(n_1)}\right]$ denote the covariance of $\underset{\sim}{M}_1^{(n_1)}$ and $\underset{\sim}{M}_2^{(n_1)}$. Since it is assumed that patient no-shows are independent of each other and the no-show rates of fixed and open appointments (γ_1 and γ_2) are known, the variance of the number of patients consulted by a provider in a session is

$$\begin{aligned}V_{n_1}(\underset{\sim}{M})&=\gamma_1(1-\gamma_1)E\left[\underset{\sim}{M}_1^{(n_1)}\right]+(1-\gamma_1)^2Var\left[\underset{\sim}{M}_1^{(n_1)}\right]+\gamma_2(1-\gamma_2)E\left[\underset{\sim}{M}_2^{(n_1)}\right]\\&\quad+(1-\gamma_2)^2Var\left[\underset{\sim}{M}_2^{(n_1)}\right]+2(1-\gamma_1)(1-\gamma_2)Cov\left[\underset{\sim}{M}_1^{(n_1)},\underset{\sim}{M}_2^{(n_1)}\right].\end{aligned} \tag{A.5}$$

When the limit of fixed appointments to be scheduled increases from n_1-1 to n_1, the number of fixed appointments scheduled increases by 1 if $\underset{\sim}{D}_1>n_1-1$, and the number of open appointments scheduled decreases by 1 if $\underset{\sim}{D}_1>n_1-1$ and $\underset{\sim}{D}_2\ge N-(n_1-1)$. Therefore, we can obtain

$$\begin{aligned}E\left[\left(\underset{\sim}{M}_1^{(n_1)}\right)^2\right]-E\left[\left(\underset{\sim}{M}_1^{(n_1-1)}\right)^2\right]&=(2n_1-1)\times P\left(\underset{\sim}{D}_1>n_1-1\right)\\&\quad+0\times P\left(\underset{\sim}{D}_1\le n_1-1\right)=(2n_1-1)[1-F_1(n_1-1)]\end{aligned} \tag{A.6}$$

$$\begin{aligned}E\left[\left(\underset{\sim}{M}_2^{(n_1)}\right)^2\right]-E\left[\left(\underset{\sim}{M}_2^{(n_1-1)}\right)^2\right]&=[-2(N-n_1)-1]\times P\left(\underset{\sim}{D}_1>n_1-1\text{and}\,\underset{\sim}{D}_2\ge N-n_1+1\right)\\&\quad+0\times P\left(\underset{\sim}{D}_1\le n_1-1\text{or}\underset{\sim}{D}_2<N-n_1+1\right)=-[2(N-n_1)\\&\quad+1][1-F_1(n_1-1)-F_2(N-n_1)+F(n_1-1,N-n_1)],\end{aligned} \tag{A.7}$$

and

$$\begin{aligned}E\left[\underset{\sim}{M}_1^{(n_1)}\underset{\sim}{M}_2^{(n_1)}\right]-E\left[\underset{\sim}{M}_1^{(n_1-1)}\underset{\sim}{M}_2^{(n_1-1)}\right]&=0\times P\left(\underset{\sim}{D}_1\le n_1-1\right)+\sum_{d_1=n_1}^{\infty}\sum_{d_2=0}^{N-n_1}d_2p(d_1,d_2)\\&\quad+[n_1(N-n_1)-(n_1-1)(N-n_1+1)]\\&\quad\times P\left(\underset{\sim}{D}_1>n_1-1\text{ and }\underset{\sim}{D}_2\ge N-n_1+1\right)\\&=(N-2n_1+1)[1-F_1(n_1-1)-F_2(N-n_1)\\&\quad+F(n_1-1,N-n_1)]+\sum_{d_1=n_1}^{\infty}\sum_{d_2=0}^{N-n_1}d_2p(d_1,d_2).\end{aligned} \tag{A.8}$$

Thus, we know

$$\begin{aligned}\mathrm{Var}\left[\underset{\sim}{M}_1^{(n_1)}\right]&=\mathrm{Var}\left[\underset{\sim}{M}_1^{(n_1-1)}\right]+\left\{E\left[\left(\underset{\sim}{M}_1^{(n_1)}\right)^2\right]-E\left[\left(\underset{\sim}{M}_1^{(n_1-1)}\right)^2\right]\right\}\\&\quad-\left\{E^2\left[\underset{\sim}{M}_1^{(n_1)}\right]-E^2\left[\underset{\sim}{M}_1^{(n_1-1)}\right]\right\}=\mathrm{Var}\left[\underset{\sim}{M}_1^{(n_1-1)}\right]\\&\quad+\left\{2n_1-1-E\left[\underset{\sim}{M}_1^{(n_1)}\right]-E\left[\underset{\sim}{M}_1^{(n_1-1)}\right]\right\}\left[1-F_1(n_1-1)\right],\end{aligned} \tag{A.9}$$

$$\begin{aligned}\mathrm{Var}\left[\underset{\sim}{M}_2^{(n_1)}\right]&=\mathrm{Var}\left[\underset{\sim}{M}_2^{(n_1-1)}\right]+\left\{E\left[\left(\underset{\sim}{M}_2^{(n_1)}\right)^2\right]-E\left[\left(\underset{\sim}{M}_2^{(n_1-1)}\right)^2\right]\right\}\\&\quad-\left\{E^2\left[\underset{\sim}{M}_2^{(n_1)}\right]-E^2\left[\underset{\sim}{M}_2^{(n_1-1)}\right]\right\}=\mathrm{Var}\left[\underset{\sim}{M}_2^{(n_1-1)}\right]\\&\quad-\left\{2(N-n_1)+1-E\left[\underset{\sim}{M}_2^{(n_1)}\right]-E\left[\underset{\sim}{M}_2^{(n_1-1)}\right]\right\}G(n_1-1,N-n_1),\end{aligned} \tag{A.10}$$

and

$$Cov\left[\underset{\sim}{M}_1^{(n_1)},\underset{\sim}{M}_2^{(n_1)}\right] = Cov\left[\underset{\sim}{M}_1^{(n_1-1)},\underset{\sim}{M}_2^{(n_1-1)}\right] + \left\{E\left[\underset{\sim}{M}_1^{(n_1)}\underset{\sim}{M}_2^{(n_1)}\right]\right.$$
$$\left.-E\left[\underset{\sim}{M}_1^{(n_1-1)}\underset{\sim}{M}_2^{(n_1-1)}\right]\right\} - \left\{E\left[\underset{\sim}{M}_1^{(n_1)}\right]E\left[\underset{\sim}{M}_2^{(n_1)}\right]\right.$$
$$\left.-E\left[\underset{\sim}{M}_1^{(n_1-1)}\right]E\left[\underset{\sim}{M}_2^{(n_1-1)}\right]\right\} = G\left(n_1-1,N-n_1\right)$$
$$\times\left\{N-2n_1+1+E\left[\underset{\sim}{M}_1^{(n_1-1)}\right]\right\}$$
$$-[1-F_1(n_1-1)]E\left[\underset{\sim}{M}_2^{(n_1)}\right]$$
$$+\sum_{d_1=n_1}^{\infty}\sum_{d_2=0}^{N-n_1} d_2 p(d_1,d_2). \quad (A.11)$$

By substituting Eqs. (A.2), (A.3), (A.9), (A.10) and (A.11) into Eq. (A.5), Eq. (6) is proven.

Next, we derive the initial condition for calculating $V_{n_1}(\underset{\sim}{M})$. When $n_1=0$, $\underset{\sim}{M}_1^{(0)}=0$ and $\underset{\sim}{M}_2^{(0)}=\min\left(\underset{\sim}{D}_2,N\right)$. Thus, we have $E\left[\underset{\sim}{M}_1^{(0)}\right]=0$, $Var\left[\underset{\sim}{M}_1^{(0)}\right]=0$, $Cov\left[\underset{\sim}{M}_1^{(0)},\underset{\sim}{M}_2^{(0)}\right]=0$, and

$$V_0(\underset{\sim}{M}) = \gamma_2(1-\gamma_2)E\left[\underset{\sim}{M}_2^{(0)}\right] + (1-\gamma_2)^2\,\mathrm{Var}\left[\underset{\sim}{M}_2^{(0)}\right]. \quad (A.12)$$

Since $\underset{\sim}{M}_2^{(0)} = \min\left(\underset{\sim}{D}_2,N\right)$, we know

$$E\left[\left(\underset{\sim}{M}_2^{(0)}\right)^2\right] = \sum_{d_2=0}^{N} d_2^2 p_2(d_2) + \sum_{d_2=N+1}^{\infty} N^2 p_2(d_2)$$
$$= N^2 - \sum_{d_2=0}^{N}\left(N^2-d_2^2\right)p_2(d_2). \quad (A.13)$$

$$\text{Thus,}\quad V\left[\underset{\sim}{M}_2^{(0)}\right] = E\left[\left(\underset{\sim}{M}_2^{(0)}\right)^2\right] - E^2\left[\underset{\sim}{M}_2^{(0)}\right] = N^2 - \sum_{d_2=0}^{N}\left(N^2-d_2^2\right)p_2(d_2)$$
$$-\left\{N-\sum_{d_2=0}^{N}(N-d_2)p_2(d_2)\right\}^2$$
$$= \sum_{d_2=0}^{N}(N-d_2)^2 p_2(d_2) - \left[\sum_{d_2=0}^{N}(N-d_2)p_2(d_2)\right]^2. \quad (A.14)$$

Substituting Eqs. (A.4) and (A.14) into Eq. (A.12), we obtain Eq. (7).

Appendix B. Procedure to find all Pareto optimal solutions to problem (P_0)

Step 1 Input the total number of appointments available, N, the no-show rates γ_1 and γ_2, and the joint probability mass function $p(d_1,d_2)$ of $\underset{\sim}{D}_1$ and $\underset{\sim}{D}_2$ for $0\le d_1\le N$ and $0\le d_2\le N$.

Step 2 Calculate $p_1(d_1)$, $p_2(d_2)$, $F_1(d_1)$, $F_2(d_2)$, and $F(d_1,d_2)$ for $0\le d_1\le N$ and $0\le d_2\le N$.

Step 3 Let $i=0$. Set $C_1(i)=0$, $C_2(i)=\sum_{d_2=0}^{N}(N-d_2)p_2(d_2)$ and $C_3(0,N-1)=\sum_{d_2=0}^{N-1}d_2[p_2(d_2)-p(0,d_2)]$. Calculate $E_i(\underset{\sim}{M})$ and $V_i(\underset{\sim}{M})$ using Eqs. (5) and (7), respectively.

Step 4 Let $i=i+1$. Calculate $G(i-1,N-i)=1-F_1(i-1)-F_2(N-i)+F(i-1,N-i)$, $C_1(i)=C_1(i-1)+F_1(i-1)$, $C_2(i)=C_2(i-1)-1+G(i-1,N-i)$, and then calculate $E_i(\underset{\sim}{M})$ and $V_i(\underset{\sim}{M})$ using Eqs. (4) and (6), respectively.

Step 5 If $i=N$, go to Step 6; otherwise, calculate $C_3(i,N-i-1)=C_3(i-1,N-i)-\sum_{d_2=0}^{N-i-1}d_2p(i,d_2)-(N-i)\left[p_2(N-i)-\sum_{d_1=0}^{i-1}p(d_1,N-i)\right]$ and go to Step 4.

Step 6 Sort n_1 in the descending order of $E_{n_1}(\underset{\sim}{M})$ into $\{n_1^{(0)}, n_1^{(1)}, \cdots, n_1^{(N)}\}$, i.e. $E_{n_1^{(0)}}(\underset{\sim}{M})\ge E_{n_1^{(1)}}(\underset{\sim}{M})\ge\cdots\ge E_{n_1^{(N)}}(\underset{\sim}{M})$.

Step 7 Let the set of the Pareto optimal solutions, S, be empty, i.e. $S=\{\}$.

Step 8 Let $i=0$. Set $S=S\cup\{n_1^{(i)}\}$.

Step 9 Let $i=i+1$. If $V_{n_1^{(i)}}(\underset{\sim}{M})<V_{n_1^{(j)}}(\underset{\sim}{M})$ for all $n_1^{(j)}\in S$, then $S=S\cup\{n_1^{(i)}\}$.

Step 10 If $i=N$, stop; otherwise, go to Step 9.

References

[1] B. Armstrong, O. Levesque, J.B. Perlin, C. Rick, G. Shectman, Reinventing veterans health administration: focus on primary care, Journal of Healthcare Management 50 (2005) 399–408.
[2] A.G. Bean, J. Talaga, Predicting appointment breaking, Journal of Health Care Marketing 15 (1995) 29–34.
[3] J. Bowers, G. Mould, Managing uncertainty in orthopedic trauma theatres, European Journal of Operational Research 154 (2004) 599–608.
[4] D.G. Bundy, G.D. Randolph, M. Murray, J. Anderson, P.A. Margolis, Open access in primary care: results of a North Carolina pilot project, Pediatrics 116 (2005) 82–87.
[5] T. Cayirli, E. Veral, Outpatient scheduling in health care: a review of literature, Production and Operations Management 12 (2003) 519–549.
[6] S. Chopra, W. Lovejoy, C. Yano, Five decades of operations management and the prospects ahead, Management Science 50 (2004) 8–14.
[7] R.W. Day, M.D. Dean, R. Garfinkel, S. Thompson, Improving patient flow in a hospital through dynamic allocation of cardiac diagnostic testing time slots, Decision Support Systems 49 (2010) 463–473.
[8] B. Denton, D. Gupta, A sequential bounding approach for optimal appointment scheduling, IIE Transactions 35 (2003) 1003–1016.
[9] B.E. Fries, V.P. Marathe, Determination of optimal variable-sized multiple-block appointment systems, Operations Research 29 (1981) 324–344.
[10] N.B. Gailmard, Short-term strategies to increase patient demand, Optometric Management (2010) www.optometricmanagement.com/mtotw/tip_new.asp?tip=445.
[11] A. George, G. Rubin, Non-attendance in general practice: a systematic review and its implications for access to primary health care, Family Practice 20 (2003) 178–184.
[12] D.C. Goodman, Twenty-year trends in regional variations in the U.S. physician workforce, Health Affairs - Web Exclusive (2004) VAR90–VAR97.
[13] D.C. Goodman, E.S. Fisher, Physician workforce crisis? Wrong diagnosis, wrong prescription, The New England Journal of Medicine 358 (2008) 1658–1661.
[14] D. Gupta, B. Denton, Appointment scheduling in health care: challenges and opportunities, IIE Transactions 40 (2008) 800–819.
[15] S. Herriott, Reducing delays and waiting times with open-office scheduling, Family Practice Management 6 (1999) 38–43.
[16] B.J. Johnson, J.W. Mold, J.M. Pontious, Reduction and management of no-shows by family medicine residency practice exemplars, Annals of Family Medicine 5 (2007) 534–539.
[17] J.G. Kennedy, J.T. Hsu, Implementation of an open access scheduling system in a residency training program, Family Medicine 35 (2003) 666–670.
[18] K.J. Klassen, T.R. Rohleder, Scheduling outpatient appointments in a dynamic environment, Journal of Operations Management 14 (1996) 83–101.
[19] L.R. LaGanga, S. Lawrence, Clinic overbooking to improve patient access and increase provider productivity, Decision Sciences 38 (2007) 251–276.
[20] M. Lamiri, F. Grimaud, X. Xie, Optimization methods for a stochastic surgery planning problem, International Journal of Production Economics 120 (2009) 400–410.
[21] S. Leader, M. Moon, Medicare trends in ambulatory surgery, Health Affairs, Web Exclusives 8 (1989) 158–170.
[22] V.J. Lee, A. Earnest, M.I. Chen, B. Krishnan, Predictors of failed attendances in a multi-specialty outpatient centre using electronic databases, BMC Health Services Research 5 (2005) 51–58.
[23] C.-J. Liao, C.D. Pegden, M. Rosenshine, Planning timely arrivals to a stochastic production or service system, IIE Transactions 25 (1993) 63–73.
[24] L. Liu, X. Liu, Block appointment systems for outpatient clinics with multiple doctors, Journal of the Operational Research Society 49 (1998) 1254–1259.
[25] L. Liu, X. Liu, Dynamic and static job allocation for multi-server systems, IIE Transactions 30 (1998) 845–854.
[26] C. Martin, T. Perfect, G. Mantle, Non-attendance in primary care: the views of patients and practices on its causes, impact, and solutions, Family Practice 22 (2005) 638–643.
[27] Massachusetts Medical Society, MMS Physician Workforce Study, Massachusetts Medical Society, Waltham, MA, 2007 [2007].

[28] D. McFadden, Economic choices, The American Economic Review 91 (2001) 351–378.
[29] A. Mehrotra, L. Keehl-Markowitz, J.Z. Ayanian, Implementing open-access scheduling of visits in primary care practices: a cautionary tale, Annals of Internal Medicine 148 (2008) 915–922.
[30] M. Murray, C. Tantau, Same-day appointments: exploding the access paradigm, Family Practice Management 7 (2000) 45–50.
[31] M. Murray, T. Bodenheimer, D. Rittenhouse, K. Grumbach, Improving timely access to primary care—case studies of the advanced access model, Journal of the American Medical Association 289 (2003) 1042–1046.
[32] M. Murray, M. Davis, B. Boushon, Panel size: how many patients can one doctor manage? Family Practice Management 14 (2007) 44–51.
[33] K. Muthuraman, M. Lawley, A stochastic overbooking model for outpatient clinical scheduling with no-shows, IIE Transactions 40 (2008) 820–837.
[34] C.D. O'Hare, J. Corlett, The outcomes of open-access scheduling, Family Practice Management 11 (2004) 35–38.
[35] D.H. Parente, M.B. Pinto, J.C. Barber, A pre-post comparison of service operational efficiency and patient satisfaction under open access scheduling, Health Care Management Review 30 (2005) 220–228.
[36] S. Pierdon, T. Charles, K. McKinley, L. Myers, Implementing advanced access in a group practice network, Family Practice Management 11 (2004) 35–38.
[37] D.J. Power, R. Sharda, Model-driven decision support systems: concepts and research directions, Decision Support Systems 43 (2007) 1044–1061.
[38] X. Qu, R.L. Rardin, J.A.S. Williams, D.R. Willis, Matching daily healthcare provider capacity to demand in advanced access scheduling systems, European Journal of Operational Research 183 (2007) 812–826.
[39] H.P. Rodriguez, W.H. Rogers, R.E. Marshall, D.G. Safran, The effects of primary care physician visit continuity on patients' experiences with care, Journal of General Internal Medicine 22 (2007) 787–793.
[40] J.E. Rohrer, M. Bernard, J. Naessens, J. Furst, K. Kircher, S. Adamson, Impact of open-access scheduling on realized access, Health Service Management Research 20 (2007) 134–139.
[41] J.W. Saultz, J. Lochner, Interpersonal continuity of care and care outcomes: a critical review, Annals of Family Medicine 3 (2005) 159–166.
[42] S.M. Schappert, E.A. Rechtsteiner, Ambulatory medical care utilization estimates for 2006, National Health Statistics Reports, National Center for Health Statistics, 2008.
[43] W. Shonick, B.W. Klein, An approach to reducing the adverse effects of broken appointments in primary care systems, Medical Care 15 (1977) 419–429.
[44] L.I. Solberg, M.C. Hroscikoski, J.M. Sperl-Hillen, P.J. O'Connor, B.F. Crabtree, Key issues in transforming health care organizations for quality: the case of advanced access, Joint Commission Journal on Quality and Safety 30 (2004) 15–24.
[45] K.E. Train, Discrete Choice Methods with Simulation, Cambridge University Press, Cambridge, UK, 2003.
[46] W.M. Valenti, J. Bookhardt-Murray, Advanced-access scheduling increases quality, productivity, and revenue, The AIDS Reader 14 (2004) 220–224.
[47] J. Vissers, Selecting a suitable appointment system in an outpatient setting, Medical Care 17 (1979) 1207–1220.
[48] E.W. Woodcock, Managing your appointment 'no-shows', The Journal of Medical Practice Management 15 (2000) 284–288.

Xiuli Qu is an assistant professor in the Department of Industrial and Systems Engineering at North Carolina Agricultural and Technical State University. Dr Qu received her PhD in industrial engineering from Purdue University. Before that, she received a BSEE and MSEE from the University of Science and Technology Beijing. Her research and teaching interests focus on health care engineering, sustainable systems, and stochastic modeling and optimization.

Ronald L. Rardin is the John and Mary Lib White Systems Integration Chair and a Distinguished Professor of Industrial Engineering at the University of Arkansas—Fayetteville. He is also the head of the University's Center on Innovation in Healthcare Logistics. Professor Rardin retired as Professor Emeritus of Industrial Engineering at Purdue University after directing the Purdue Energy Modeling Research Groups and playing a leading role in its Regenstrief Center for Healthcare Engineering. He also served a rotation from 2000 to 2003 as Program Director for Operations Research and Service Enterprise Engineering at the National Science Foundation including founding the program to foster research in service industries, especially healthcare. Dr. Rardin obtained his B.A. and M.P.A. degrees from the University of Kansas, and after working in city government, consulting and distribution for five years, a Ph.D. at Georgia Institute of Technology. His current teaching and research interests center on large-scale optimization modeling and algorithms, especially their applications in healthcare delivery. He is an award winning teacher of those topics, and co-author of numerous research papers and two comprehensive textbooks: a graduate text Discrete Optimization, published in 1988, and a comprehensive undergraduate textbook Optimization in Operations Research in 1998. He has been selected as a fellow of the Institute for Operations Research and the Management Sciences.

Julie Ann Stuart Williams is an associate professor in the Dept. of Management and MIS at the University of West Florida. Professor Williams received her PhD in industrial and systems engineering from Georgia Institute of Technology. Before moving to UWF, she was a faculty member at Purdue University and The Ohio State University. She is a registered professional engineer in the state of Ohio. Her research and teaching interests focus on sustainable operations and health care systems.