Operations Research for Health Care 6 (2015) 1–10

Contents lists available at ScienceDirect

Operations Research for Health Care

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

An overbooking scheduling model for outpatient appointments in a multi-provider clinic

Moh'd El-Sharo, Bichen Zheng, Sang Won Yoon *, Mohammad T. Khasawneh

Department of Systems Science and Industrial Engineering, State University of New York at Binghamton, Binghamton, NY 13902, United States

ARTICLE INFO

Article history:
Received 27 March 2014
Accepted 13 May 2015
Available online 24 June 2015

Keywords:
Patient appointment scheduling
Overbooking model
Multiple provider clinics

ABSTRACT

In general, most outpatient clinics are staffed by more than one provider in order to satisfy high patient demand and distribute the overall workload. Although several overbooking scheduling models have been proposed in the literature for single-provider settings, it would be difficult to extend and apply them directly to a multi-provider setting. This research proposes an overbooking scheduling model for multiple-provider clinics to optimize the number of overbooked patients and maximize the expected profit. In addition, the proposed model incorporates various probabilities of patient no-shows to simulate actual outpatient clinic characteristics. Experimental results indicate that the proposed multi-provider overbooking model outperforms dissociated single-provider overbooking models in terms of increasing the expected profit by 19–24%, and the number of scheduled patients by 4.5–9%. The results also indicate that the expected profits will be maximized by redistributing overflowing patients based on providers' workloads when varying the clinic schedule capacity, number of providers, and patient no-show probability.

1. Introduction

Outpatient services have grown in multiple capacities during the last two decades in terms of their volume, technology, and variety. For instance, estimates from the late 1990s indicate that 60% of all surgeries undertaken in the UK and 79% in Denmark are same-day surgeries, but these are still lower than in Canada (85%) and the USA (94%) [1]. Not only the same-day surgery, but also there are a number of outpatient services provided at the hospital, including blood transfusions, laboratory test, radiation services, dental treatment, preventive and screening services, etc. These short-term appointments are typically fully occupied when patient volume is close to a clinic's capacity. This leads to many patients having to wait for extended periods of time, and as a result increases the possibility of missing appointments, health condition attenuation, and even death [2]. In addition, it has been reported that the effectiveness of patient appointment scheduling is highly correlated with patient satisfaction and hospital revenue. Therefore, it is important to have an effective outpatient appointment scheduling method, that reduces patient waiting time, increases medical staff and equipment utilization, and reduces healthcare facilities' operating costs [3]. Outpatient scheduling is one of the most challenging problems for healthcare providers, primarily due to uncertain patient no-shows and cancellations [4].

Healthcare outpatient clinics strive to increase their profit by maximizing their resource utilization. However, it is difficult to predict the number of patient arrivals and therefore to execute appropriate staffing and resource utilization plans when patient no-shows and cancellations are common. To overcome this obstacle, an overbooking scheduling model has been proposed in the literature [3]. Overbooking is an essential technique used to reduce the effects of patient no-shows and cancellations. For many years, it has been utilized to increase resource utilization in the airline industry [5]. Both the airline industry and healthcare services share the same consequences that passengers or patient no-shows and cancellations have on their profit. As a result, overbooking scheduling has become a considerable option for outpatient clinics [2,6]. Healthcare providers tend to overbook patient's appointments in response to patient no-shows and cancellations, a method that can be detrimental because excessive overbooking can potentially increase patient waiting time [7]. A sequential probabilistic overbooking model for single-provider outpatient clinics has been proposed to mitigate the effect of patient no-shows with variable no-show probabilities, as shown in Fig. 1 [2].

Generally, the overbooking appointment scheduling model aims to schedule more than one patient to an appointment slot, given the possibility that all scheduled patients will show up

* Corresponding author.
E-mail address: yoons@binghamton.edu (S.W. Yoon).

http://dx.doi.org/10.1016/j.orhc.2015.05.004
2211-6923/

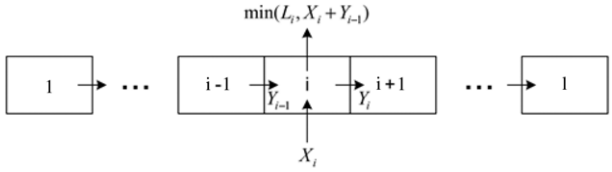

Fig. 1. An overbooking model [2].

to their scheduled appointments. In this case, not all scheduled patients can be served in their appointment slots, so some of the patients have to be treated in subsequent appointment slots. To maximize the expected profit of an outpatient clinic, it is critical to optimize the number of overbooking patients, considering the costs of overflowing patients and patient no-show probability. Outpatient clinics typically record patient scheduling information including their no-shows and cancellations, which can be utilized to estimate patient no-show probability.

In this research, the patient no-show probability is modeled on the level of individual patients, as to adopt a dynamic of independent patient no-shows and cancellations [8,9]. In addition, many of the outpatient clinics are typically operated by more than one provider (i.e., multiple providers). Although those outpatient clinics are multiple-provider clinics, normally, patients prefer being scheduled with their specific providers for continuity in their care (e.g., primary care physician, dentist, or specialty care physician). The trade-offs between timely access health service and patient–physician continuity have been discussed in [10]. In their research, a generalizable analytical algorithm is developed to allocate physician resources on pre-schedule and urgent patients. The impact of combining the two cases was further studied in 2013 [11]. In this research it is assumed that patients will not have a preference to specific providers for simplicity, which allows all available providers to provide them with the necessary medical services. With these considerations, this research proposes an overbooking appointment scheduling model, one which has a fixed appointment duration, and scrutinizes the costs associated with overflowing patient waiting time, provider overtime, while also looking at patient walk-ins, variable no-show probabilities, and providers' variable service time. Ultimately, the objective of the model is to assist multi-provider outpatient clinics in maximizing their net profit.

The remainder of this article is organized as follows: the background of this research is provided in Section 2. The proposed multi-provider appointment overbooking (MPAO) model and overbooking algorithm are discussed in Section 3. The experimental results and analyses are addressed in Section 4. Finally, conclusions and future work are presented in Section 5.

2. Background of research

Many studies have been conducted to analyze patient appointment scheduling models within a variety of healthcare practices, including primary care, specialty care, and elective surgery [3,4,12–16]. An overview of the different appointment scheduling models is summarized as follows:

- Primary care appointment scheduling aims to schedule appointments with primary care physicians based on a patient's request if he/she has specific healthcare concerns. Patients typically schedule an individual appointment slot with a fixed interval [17,18].
- Specialty clinic appointment scheduling aims to schedule a specialty consultation appointment when patients are referred by their primary care physicians for specific treatments. Appointments are typically made in batches, and patients are scheduled for individual appointment slots with variable intervals [19].

These appointment scheduling models vary in terms of patient access rules, appointment start time, and patient/provider preference, which are summarized in Fig. 2. In addition, there are three main appointment scheduling policies, such as unit-level, periodic, and single batch. The unit-level scheduling is generally applied when appointment requests occur one at a time at random time intervals, and can be applied in open access scheduling models [17,18] or regular First-Call-First-Scheduled (FCFS) models [2,20,21]. Periodic scheduling is applied when a group of patient appointments accumulate and are sent to the scheduling system all at once. Single batch scheduling is applied when the schedule of appointments is not decided until all appointment requests are received [4].

Overbooking appointment scheduling has been studied with the following objective functions:

1. Maximizing monetary value, cost reduction, and profit maximization [20,22,23].
2. Reducing/eliminating patient waiting time, provider idle time, provider overtime [2,13,18,20,21].
3. Reducing the number of overflowing patients from one appointment to another as proposed in [2,20,21].

Discrete event simulation was introduced to study the outpatient clinic schedule by [24]. In their study, a Monte Carlo simulation model was developed to show the effects of alternative scheduling appoint strategies with the predicted walk-in patient arrival patterns for an individual health care provider. It is noticed that discrete event simulation can be built as a risk-free platform to test and visualize different operating strategies of a clinic [25]. To address the urgent patient, a simulation model is designed to search the best scheduling rule and the best placement of appointment slots left for urgent patients [26]. Later more studies assumed various numbers of providers; single [12] and multiple [13] providers. Their simulation models included a process of single and multiple stages, which were very specific to their situations, and cannot be generalized to all systems. A universal "Dome" appointment rule was developed considering environmental factors such as no-shows, walk-ins, number of appointments per session, variability of service times, and cost of doctor's time to patients' time. It performed well with a low total system cost based on the simulation results [27]. Alternative analytical models were also developed for single-provider clinics [2,6,14–16,21,28–32]. A simulation based optimization model was used to determine which particular appointment slots to double-book or leave open with the consideration of seasonal walk-in rates [33]. Although the various arrival rates of walk-in patients were of concern in this research, no-shows were not considered. In addition, for a simplicity purpose, the clinic was assumed to be a single healthcare provider clinic. While most multiple-provider clinics can be modeled using a group of single-provider models, it is still necessary for multiple-provider clinics to study a multiple-provider model. The details of overbooking scheduling literature are summarized in Table 1.

Many researchers applied discrete event simulation to model their studies. In their models, they adapted different distributions, such as uniform, exponential, truncated Erlang, Gamma, and empirical distributions. In analytical models, most researchers used exponential and uniform distributions for the model simplicity. However, there was limited research that analyzed an overbooking scheduling model for multi-provider settings. Also, even though the walk-in patient arrival and no-shows are mentioned in the literature, there is still lacking of a comprehensive model incorporating these two important factors. In the real practice, these two types of patients are difficult to avoid but significantly affect the appointment scheduling system performance and the expected clinic profit.

In this research, a multiple provider appointment overbooking (MPAO) model is proposed to minimize patient waiting time and

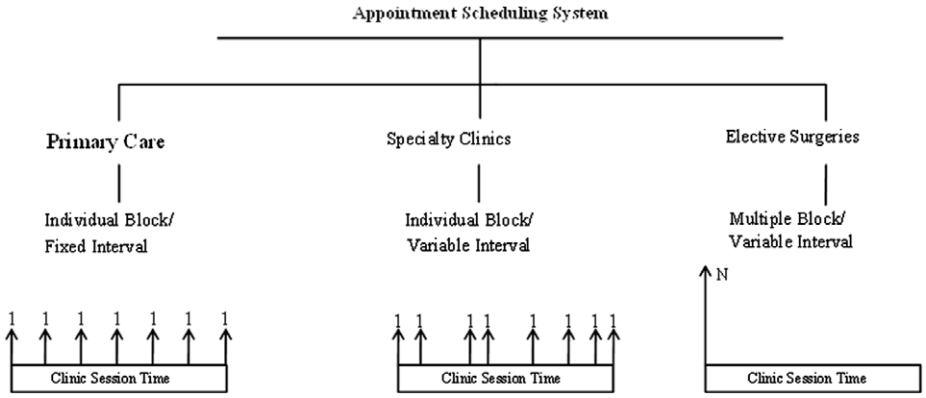

Fig. 2. Illustration of different outpatient appointment scheduling models [4].

Table 1
Overview of overbooking scheduling literature.

Research	Providers	Service distribution	Process steps	Performance measures	Methodology
Vanden Bosch and Dietz [12]	Multiple	Truncated Erlang and gamma based on empirical distributions	Single	Patient waiting time and doctor overtime	Discrete event simulation
Rohleder and Klassen [13]	Multiple	Exponential	Multiple	Overtime, overload and patient delay	Discrete event simulation
Gupta and Denton [4]	Single	Exponential	Single	Patient indirect waiting time	Two-stage stochastic linear programming
Cayirli et al. [3]	Single	Uniform	Single	Patient waiting time and providers idle time	Discrete event simulation
Kopach et al. [14]	Single	Exponential	Multiple	Fraction of patients using the open access policy, scheduling horizon, patient preference of care group, patient waiting time, and provider idle time and tardiness	Discrete event simulation and design of experiments
Kaandorp and Koole [28]	Single	Exponential	Single	Patient waiting time, provider idle time and tardiness	Probabilistic dynamic programming model and local search algorithm
Muthuraman and Lawley [2]	Single	Exponential	Single	Patient waiting time and provider idle time	Sequential probabilistic model and greedy algorithm
Patrick et al. [6]	Single	Uniform	Single	Patient waiting time and cost of operation	Markov decision process and dynamic programming
Liu et al. [29]	Single	Constant	–	Patient waiting time, provider idle time and provider overtime costs	Markov decision process and heuristic dynamic policy
Chakraborty et al. [21]	Single	General	Single	Patient waiting time and provider idle time	Sequential probabilistic model with Greedy Algorithm
Sun et al. [30]	Multiple	Exponential	Multiple	Number of served patients, patient waiting time and resource utilization	Discrete event simulation
Liao et al. [15]	Single	Exponential or Erlang-k	Single	Patient waiting time and server utilization	Dynamic programming and branch-and-bound algorithm
Erdogan and Denton [16]	Single	General	Single	Cost associated with patient waiting time	Stochastic linear programming algorithms
Cayirli et al. [33]	Single	Lognormal	Single	Patient waiting time, physician idle and overtime	Simulation based optimization

provider idle, while simultaneously maximizing clinic's profit. The no-shows and the arrival rates of walk-in patients are considered to further increase the clinic profit. Additionally, the proposed model considers a fixed appointment duration for all providers (e.g., corresponding to primary care appointments) and variable appointment durations following an exponential distribution (e.g., corresponding to a wide variety of specialized outpatient services).

3. Methodology

In this research, the multi-provider appointment overbooking (MPAO) model is proposed to maximize the expected profit of a multiple-provider clinic by mitigating the effect of patient no-shows through an overbooking scheduling approach. To develop the expected profit function in the MPAO model, various financial elements are considered, which are primarily derived from the number of patient arrivals (i.e., scheduled and walk-in patients) to each appointment slot, the number of patients served at their scheduled appointments, and the number of overflowing patients. In addition, because a long patient waiting time influences the possibility that a patient could leave without being seen, which can cause patient dissatisfaction, it is considered as an opportunity cost. Depending on the type of services, the penalty of patient waiting time can vary. In this research, an analytical model is proposed to optimize the number of overbooked patients to maximize the expected net profit. Two approaches to determine overflowing patient allocation are also studied assuming that only if a patient becomes an overflowing patient, he/she can be served by any other available provider in subsequent appointment slots; i.e., evenly (ED) vs. weighted-mean (WM) distributed patient allocations.

To simplify the overbooking systems, several major assumptions have been made as follows:

Table 2
List of mathematical notations.

Symbol	Description
i	Index for appointment slot
j	Index for each provider
$f(RD_i)$	Number of redistributed patients after appointment slot i
n	Index of patient to be scheduled when the patient call
$t_{i,j}$	Appointment time duration for appointment slot i for provider j
$A_{i,j}$	Expected number of all patient arrivals for appointment slot i for provider j
C_{OF}	Penalty cost of overflow per patient
C_{OT}	Penalty cost of overtime per patient
I	Total number of appointment slots in a day per provider
J	Total number of providers in a healthcare setting
N	Total number of patients calling for appointments in a day
$OF_{i,j}$	Number of patients overflowing from appointment slot i for provider j
OT_j	Number of patients overflow at the end of the day for provider j
R	Revenue rewarded for each patient served
RD_i	Sum of all patients overflowing from appointment i
$S_{i,j}$	Number of scheduled patients at appointment slot i for provider j
$V(\psi)$	Expected profit function
Y	Daily initial operational cost
$\alpha_{i,j}$	Probability of a patient attendance to appointment slot i for provider j
$\lambda_{i,j}$	Walk-in arrival rate per appointment duration $t_{i,j}$
ψ	Matrix of decision variables S_{ij} in the overbooking model

- When a patient becomes an overflowing patient, the patient will be assigned to any available provider to minimize the waiting time.
- There is no difference among different providers in terms of the quality of care.
- One patient can be scheduled at only one appointment slot.
- Patients are punctual if they will show-up.

A list of the mathematical notations is summarized in Table 2.

3.1. Decision variables

The outpatient appointment schedule

$$\psi = \begin{bmatrix} S_{1,1} & S_{1,2} & \cdots & S_{1,j} \\ S_{2,1} & S_{2,2} & \cdots & S_{2,j} \\ \vdots & \vdots & \ddots & \vdots \\ S_{i,1} & S_{i,2} & \cdots & S_{i,j} \end{bmatrix},$$

is optimized to maximize the expected profit, where S_{ij} is defined as the decision variable representing number of scheduled patients at appointment slot i for health-care provider j. One patient can be scheduled at most one appointment slot. For each appointment slot, the no-show probability of each scheduled patient is identical. Patient arrivals vary depending on no-show distribution at each appointment slot of a provider, and patient overbooking can occur when more than one patient is assigned for the same appointment slot. Consequently, patient arrival estimation is critical in calculating and maximizing the expected net profit.

The expected number of all patient arrivals for appointment slot i for provider j, $A_{i,j}$, should be the sum of scheduled, $S_{i,j}$ and walk-in, $W_{i,j}$, patients. Since only one patient can be served in appointment i, the number of unfulfilled patient treatments during appointment slot i overflows to appointment slot $i+1$. The number of overflowing patients at appointment slot i for provider j, $OF_{i,j}$, is defined as

$$OF_{i,j} = \max(OF_{i-1,j} + S_{i,j} + W_{i,j} - 1, 0) \tag{1}$$

$\forall OF_{i,j}$ is reallocated to any one of the available providers in appointment $i+1$, which will be discussed later in this section. There are two types of patients considered in this research; i.e., walk-in and scheduled patients. It is assumed that the overflowing patients are treated as scheduled patients without having prioritizing rules in appointment slot $i+1$. Furthermore, newly arrived patients and overflowing patients are handled identically in the proposed model. Only one patient from the pool of arrivals/overflows is treated at a time.

To estimate the net profit associated with served patients, $S_{i,j}$ and $OF_{i,j}$, $\forall i, j$ is estimated. The probability that a patient will show up to his/her appointment slot, which may vary from one appointment slot at a provider to another, is denoted by α. Therefore, because of the property of a series of Bernoulli trials, $E[S_{i,j}]$ follows a binomial distribution with $\alpha_{i,j}$.

To estimate $E[W_{i,j}]$, let $t_{i,j}$ be the time duration for appointment slot i for provider j. Let $W_{i,j}$ follow a Poisson distribution with the arrival rate λ. Then, $E[W_{i,j}] = \lambda \cdot t_{i,j}$. As a result, the total number of $S_{i,j}$ and $W_{i,j}$ can be defined as

$$A_{i,j} = S_{i,j} \cdot \alpha_{i,j} + \lambda \cdot t_{i,j} \tag{2}$$

$S_{i,j}$ and $OF_{i,j}$, $\forall i, j$ contribute to providers' workload, so that patient reallocation for $OF_{i,j}$, $\forall i, j$ between providers is considered in order to minimize the patient waiting time; i.e., $OF_{i,j}$, $\forall i, j$ can be reallocated among available providers in appointment slot i. The two patient reallocation methods are proposed in this research are:

- Evenly-distributed (ED) patient reallocation: The mean of overflowing patients from appointment slot $i-1$, $\forall j$ is allocated evenly among available providers in appointment slot i as defined by

$$f_{i,j}(RD_{i-1}) = \left\| \frac{1}{J} RD_{i-1} \right\| = \left\| \frac{1}{J} \sum_{j=1}^{J} OF_{i-1,j} \right\| \tag{3}$$

  where $RD_{i-1} = \sum_{j=1}^{J} OF_{i-1,j}$ and $f_{i,j}(RD_{i-1})$ represents the expected number of $OF_{i,j}$ as a function of RD_{i-1}.
- Weighted-mean (WM) patient reallocation: RD_{i-1} is redistributed based on providers' workload, such as $S_{i,j}$, which is defined as

$$f_{i,j}(RD_{i-1}) = \left\| \left(1 - \frac{S_{ij}}{\sum_{j=1}^{J} S_{ij}} \right) RD_{i-1} \right\| = \left\| \left(1 - \frac{S_{ij}}{\sum_{j=1}^{J} S_{ij}} \right) \sum_{j=1}^{J} OF_{i-1,j} \right\|. \tag{4}$$

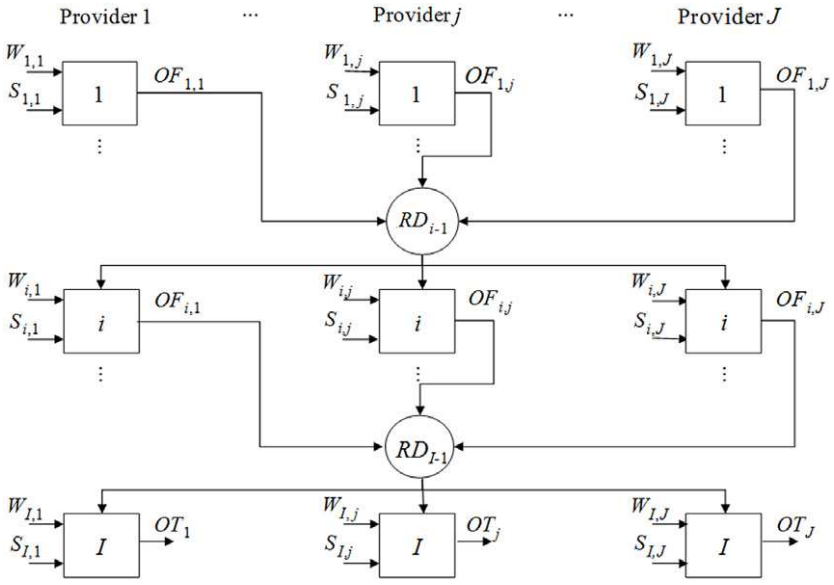

Fig. 3. Multi-provider appointment scheduling system with workload balancing.

The patient reallocation methods for a multi-provider clinic are illustrated in Fig. 3.

3.2. The expected profit function

The expected profit function can be calculated based on $E[S_{i,j}]$, $E[W_{i,j}]$, and $f_{i,j}(RD_{i-1})$ as well as overflowing patients based on $RD_I = \sum_{j=1}^{J} OF_{I,j}$. The decision variable ψ is the matrix of scheduled patients that should be determined to maximize the expected profit. The expected profit function, $V(\psi)$, is defined as

$$\begin{aligned} V(\psi) = & \sum_{i=1}^{I}\sum_{j=1}^{J} R \min\{S_{i,j}\cdot\alpha_{i,j} + \lambda\cdot t_{i,j} + f_{i,j}(RD_{i-1}),\, 1\} \\ & - \sum_{i=1}^{I}\sum_{j=1}^{J} C_{OF} \max\{S_{i,j}\cdot\alpha_{i,j} \\ & + \lambda\cdot t_{i,j} + f_{i,j}(RD_{i-1}) - 1,\, 0\} \\ & - \sum_{j=1}^{J} C_{OT} \max\{S_{I,j}\cdot\alpha_{I,j} + \lambda\cdot t_{I,j} \\ & + f_{I,j}(RD_{I-1}) - 1,\, 0\} - Y \end{aligned} \tag{5}$$

where R denotes the revenue of a patient served at each appointment, CO_{OF} and CO_{OT} denote the cost of patient overflowing and session overtime respectively, and Y denotes the clinic's initial operational cost. $V(\psi)$ is calculated by the total revenue obtained from the number of served patients minus the expected costs of overflowing (waiting) and overtime patients, and the fixed cost of operation. $f_{i,j}(RD_{i-1})$ denotes patient reallocation of overflowing patients when determining $V(\psi)$. As a result, two MPAO models (i.e., MPAO-ED and -WM models) have been developed.

In this research, patients are assigned to their appointment slots sequentially, which indicates that after each patient placement, the overall patient schedule will be different in terms of the number of patients scheduled, their overflow probabilities, and the expected profit function. When determining an appointment, each patient is assigned to an appointment slot, which maximizes the expected total profit of a clinic. This problem is a multiple variable sequential search problem. To allocate a patient in the appointment slot and identify the most appropriate slot, a search algorithm (steepest descent) is applied. Whenever a new patient appointment request arrives, the steepest descent algorithm is used to find the optimal appointment slot and provider based on the existing clinic schedule. The gradient direction based on the current schedule is calculated and the steepest descent direction is found accordingly, which determines which slot the new request should be placed on the schedule to maximize the expected profit and reduce the patient probability of overflow. In this case, the previous scheduled patients are reserved. The patient is informed of the appointment time and provider before the end of each call and the appointment would not be affected (altered or cancelled) by subsequent patient requests.

The first derivative of $V(\psi)$ is calculated in the form of a gradient vector as

$$\nabla V(\psi) = \left[\frac{\partial V}{\partial S_{1,1}}, \frac{\partial V}{\partial S_{1,2}}, \frac{\partial V}{\partial S_{1,3}}, \ldots, \frac{\partial V}{\partial S_{I,J}}\right]^T. \tag{6}$$

Eq. (6) is used to determine the direction of points (appointments) that maximizes the expected profit function. Given $\nabla V(\psi)$ exists, the gradient vector point is the direction of the greatest rate of increase of $V(\psi)$. At any initial point (appointment) $S_{0,0}$, $\nabla V(\psi)$ is normal to the contour whose value of $V(\psi)$ is constant that passes through the point $S_{0,0}$. The positive gradient point is a direction of the greatest rate of increase of $V(\psi)$ [34]. The steps followed in the steepest descent search are:

Step 1: Start at random initial points
Step 2: For $i = 1 : I$ and $i = 1 : J$, calculate $\nabla V(\psi)$
Step 3: Move in the (positive) direction of $\nabla V(\psi)$ by a unit step
Step 4: After moving to the new point, calculate the profit function value $V^+(\psi)$ of the new point
Step 5: Terminate the calculation if $|V(\psi) - V^+(\psi)| < \varepsilon$, where ε is a pre-assigned error tolerance; else, $|V(\psi) - V^+(\psi)| \geq \varepsilon$, then return to Step 2.

The pre-assigned value of ε controls the precision of determining $V^*(\psi)$. The smaller the value of ε, the more precise the search will be in its approach to the local maximum and vice versa. Fig. 4 demonstrates the sequence of steps taken to find the optimal appointment allocation, using the steepest descent algorithm.

The steepest descent algorithm starts with the schedule prior to double booking, as shown in Fig. 4(a). The algorithm generates a set of initial points $S_{0,0}$, adding the set of initial points to the schedule as shown in Fig. 4(b). Then, it searches for a greater value

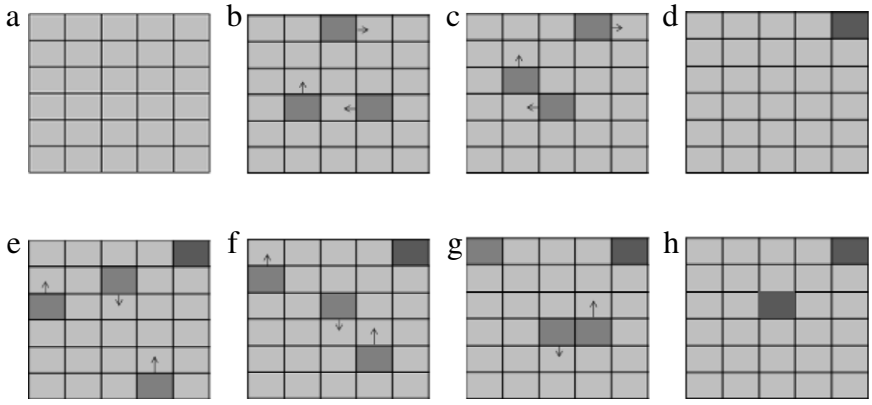

Fig. 4. Illustration of patient overbooking appointment allocation sequence (i.e., (a)–(h)) using the steepest descent algorithm.

of $V(\psi^n)$ by changing $S_{0,0}$ using the steepest descent as illustrated in Fig. 4(b)–(d). The algorithm finds the optimal allocation that leads to $V^*(\psi)$ as shown in Fig. 4(d). The same steps are repeated for each patient. Fig. 4(e)–(h) shows the steepest descent algorithm steps used to schedule the subsequent patient. Then, a greedy algorithm is used to identify the stopping criterion (maximum number of patients) that maximizes the clinic's expected profit.

3.3. Overbooking appointment scheduling

To attain the optimal number of overbooked patients, a balance between resource utilization and patient waiting time should be considered. When a patient does not come to his/her appointment time and the designated provider has no other arrivals, the provider is idle and the clinic loses the revenue expected from treating the patient in that appointment. Scheduling more than one patient per appointment slot anticipates patient no show, and increases the probability that providers are utilized. However, a penalty of excessive overbooking may be incurred, such as longer patient waiting time and overflowing patients to later appointment slots. Also, it may cause an increase in patient waiting time and in patient overflow from one appointment to another. The overbooking scheduling problem can be easily defined as a sequential problem. Therefore, $V(\psi)$ must be maximized each time a patient is placed in the appointment slot. The proposed scheduling algorithm can be explained as a set of iterations that find the optimal appointment location in the schedule for each calling patient, which is illustrated in detail in Algorithm 1. The stopping criterion is when the Nth patient is scheduled and scheduling one or more patients exceeding the Nth patient will result in less expected profit than with N patients.

3.4. Numerical examples

In a multi-provider outpatient clinic, a typical schedule table is shown in Table 3, which is generated by the Round-Robin (RR) [2], the Sequential Scheduling Method (SSM)[2], the proposed MPAO-ED and MPAO-WM algorithms. Each appointment slot can be scheduled more than one appointments, but one patient can be only assigned to one appointment slot ($i = 1, 2, 3$). In this numerical example, there is a total of 10 patients used for scheduling appointments, two providers ($j = 1, 2$), and three time slots. The RR is to assign the kth patient to slot $((k-1) \bmod 3)+1$ [2]. Assuming that the no-show probabilities are given, the schedule can be obtained by the SSM. The SSM algorithm calculates the arrival probability and overflow probability matrix first. By maximizing the expected profit, the optimal schedule can be found with the consideration of no-show and overflow probability. Based on the numerical example, the SSM schedules more patients (i.e., patient 1, 7, 9) at the same appointment slot in case some of patients would not show up. Based on the SSM generated schedule, a clinical environment with walk-in patients is used for validating the performance of the SSM. Compared to the RR and SSM, the proposed MPAO-ED and MPAO-WM algorithms consider both no-show rates and walk-in patients in the analytical model before the schedule is fixed. As shown in Tables 3 and 4, more appointment slots can be flexible to accommodate potential walk-in patients using MPAO-ED and MPAO-WM (e.g., $i = 2$ in Table 3 and $i = 4$ in Table 4). Since the arrival times of walk-in patients are difficult to estimate, all appointment slots can be utilized to undertake walk-in patients. However, to balance the scheduled and walk-in patients, certain adjustments have been made based on MPAO-ED and -WM

Algorithm 1: The MPAO scheduling algorithm

select I, J // Total number of appointments/day and providers
initialize
$S_{i,j} = 0, n = 0$ // Initialize decision variables
$V(\psi^0) = 0$ // Initialize the expected profit with empty schedule
Wait for a patient appointment request
repeat
  $n = n + 1$ // Increase patient count by one

$$V(\psi^n) = \sum_{i=1}^{I}\sum_{j=1}^{J} R\min\{S_{i,j}\cdot\alpha_{i,j} + \lambda\cdot t_{i,j} + f_{i,j}(RD_{i-1}), 1\} - \sum_{i=1}^{I}\sum_{j=1}^{J} C_{OF}\max\{S_{i,j}\cdot\alpha_{i,j} + \lambda\cdot t_{i,j} + f_{i,j}(RD_{i-1}) - 1, 0\} - \sum_{j=1}^{J} C_{OT}\max\{S_{I,j}\cdot\alpha_{I,j} + \lambda\cdot t_{I,j} + f_{I,j}(RD_{I-1}) - 1, 0\} - Y$$

  // Compute the profit function
  $\nabla V(\psi) = \left[\frac{\partial V}{\partial S_{1,1}}, \frac{\partial V}{\partial S_{1,2}}, \frac{\partial V}{\partial S_{1,3}}, \cdots, \frac{\partial V}{\partial S_{I,J}}\right]^T$
  Use steepest descent algorithm to maximize $V^n(\psi)$
  $(i,j)^* = \arg_{(i,j)} \max V(\psi^n)$ // Identify index as new sub-optimal
  $S_{(i,j)^*} = S_{(i,j)^*} + 1$
until $V(\psi^n) < V(\psi^{n-1})$
$N = n - 1$ // Maximum number of scheduled patient
return ψ^N // Return the best schedule found

Table 3
A numerical example of a clinic scheduling table generated by four algorithms.

	Patient index							
	RR		SSM		MPAO-ED		MPAO-WM	
	$j=1$	$j=2$	$j=1$	$j=2$	$j=1$	$j=2$	$j=1$	$j=2$
$i=1$	1, 7	2, 8	1, 7, 9	2, 8, 10	1, 7	2, 8	1, 7, 8	2, 9
$i=2$	3, 9	4, 10	3	4	3	4	3	4
$i=3$	5	6	5	6	5, 9	6, 10	5, 10	6

Table 4
A realistic implementation of scheduling 70 patient in a clinic with two health care providers ($\alpha = 10\%$ and $\lambda = 0.5$).

	Scheduled patient quantity							
	RR		SSM		MPAO-ED		MPAO-WM	
	$j=1$	$j=2$	$j=1$	$j=2$	$j=1$	$j=2$	$j=1$	$j=2$
$i=1$	3	3	3	3	3	2	3	1
$i=2$	3	3	1	2	3	2	4	1
$i=3$	3	3	3	3	3	1	4	2
$i=4$	2	2	3	3	4	1	3	1
$i=5$	2	2	1	2	3	1	3	1
$i=6$	2	2	3	3	3	1	4	1
$i=7$	2	2	3	2	3	1	4	1
$i=8$	2	2	3	3	4	1	3	1
$i=9$	2	2	3	3	4	1	3	1
$i=10$	2	2	2	2	3	1	3	1
$i=11$	2	2	2	2	3	1	3	1
$i=12$	2	2	1	1	3	1	3	1
$i=13$	2	2	2	2	3	0	3	1
$i=14$	2	2	2	2	4	0	3	1
$i=15$	2	2	3	2	3	0	3	1
$i=16$	2	2	0	0	4	3	3	2

Table 5
A summary of parameter settings.

Parameters	Values
Number of appointment slots	$I = [7, 45]$
Number of providers	$J = [2, 10]$
Revenue per visit	$R = \$100$
Cost of overflow	$C_{OF} = \$25$
Cost of overtime	$C_{OT} = \$25$
Attendance probability	$\alpha = 0.1, 0.5, 0.9$
Walk-in rate per appointment	$\lambda = 0.1$
Error tolerance of steepest descent	$\varepsilon = 0.001$

according to no-show rates and weighted healthcare providers. From MPAO-ED scheduling results in Table 4, for instance, the health-care provider $j = 2$ is under capacity for scheduling due to the walk-in patients in the appointment slot $i = 3$. Generally speaking, the proposed MPAO algorithm schedules fewer patients to make more appointment slots available for walk-in patients and overflow patients from previous appointment slots. The expected profit function including patient treatment reward, overflowing cost and overtime cost are maximized using steepest descent algorithm.

4. Experimental results and analyses

In this research, the MPAO model has been proposed when overflowing patients can be treated by any other providers. To validate the effectiveness of the proposed MPAO model, a set of numerical experiments have been conducted with different system parameters, which is summarized in Table 5.

The proposed MPAO models have been compared to the RR and the SSM [2] in terms of the number of patients scheduled for all providers, the expected profit, and the number of overflowing patients. Single provider models (the RR and the SSM) were simulated for multiple providers by repeating simulation multiple times and compared to the MPAO with the same number of providers (e.g., the SSM and the MPAO were simulated for two providers in each setting and results were compared). Incoming patients are evenly re-distributed to different providers. The RR method is based on a naive sequential method of scheduling patients in appointments sequentially regardless of any performance measure. On the other hand, the SSM, which maximizes the expected profit, places patients in a schedule sequentially for a single provider. Although the RR and the SSM models do not consider walk-in patients, for comparison purposes, walk-ins with equivalent Poisson distributions were simulated in each experimental run. The overall experimental workflow is illustrated in Fig. 5.

4.1. Analysis of total expected profit

The MPAO-ED and the MPAO-WM model have been analyzed in terms of the expected profits when $S_{i,j}$ is varied. The simulation results are illustrated in Fig. 6. The results indicate that the MPAO models increase $V(\psi)$ by 19% and 24%, for the SSM and RR methods respectively. In addition, $V(\psi)$ of the MPAO-WM model is 5% greater than of the MPAO-ED method.

As shown in Fig. 6, the $V^*(\psi)$ obtained with the MPAO models are higher than the RR and SSM methods by 4.5% and 9%, respectively. This implies that if outpatient clinics apply the proposed MPAO model, then they can serve more patients with a lower probability of having overflowing patients. Moreover, $V(\psi)$ is negative when the total cost of overflow and overtime exceeds the total revenue, depending on Y, R, and $1 - \alpha$. It can be seen that $V(\psi)$ increases linearly when $S_{i,j} < I$, since only I patient appointments are scheduled for each provider; i.e., $OF_{i,j} = 0$.

4.2. Analysis of overflowing patients

The MPAO models have also been analyzed in terms of the number of overflowing patients. The MPAO-WM model yields the lowest number of overflowing patients, which maximizes the expected profit, as shown in Fig. 7. When the number of scheduled patients is below clinic capacity, $OF_{i,j} = 0$; therefore,

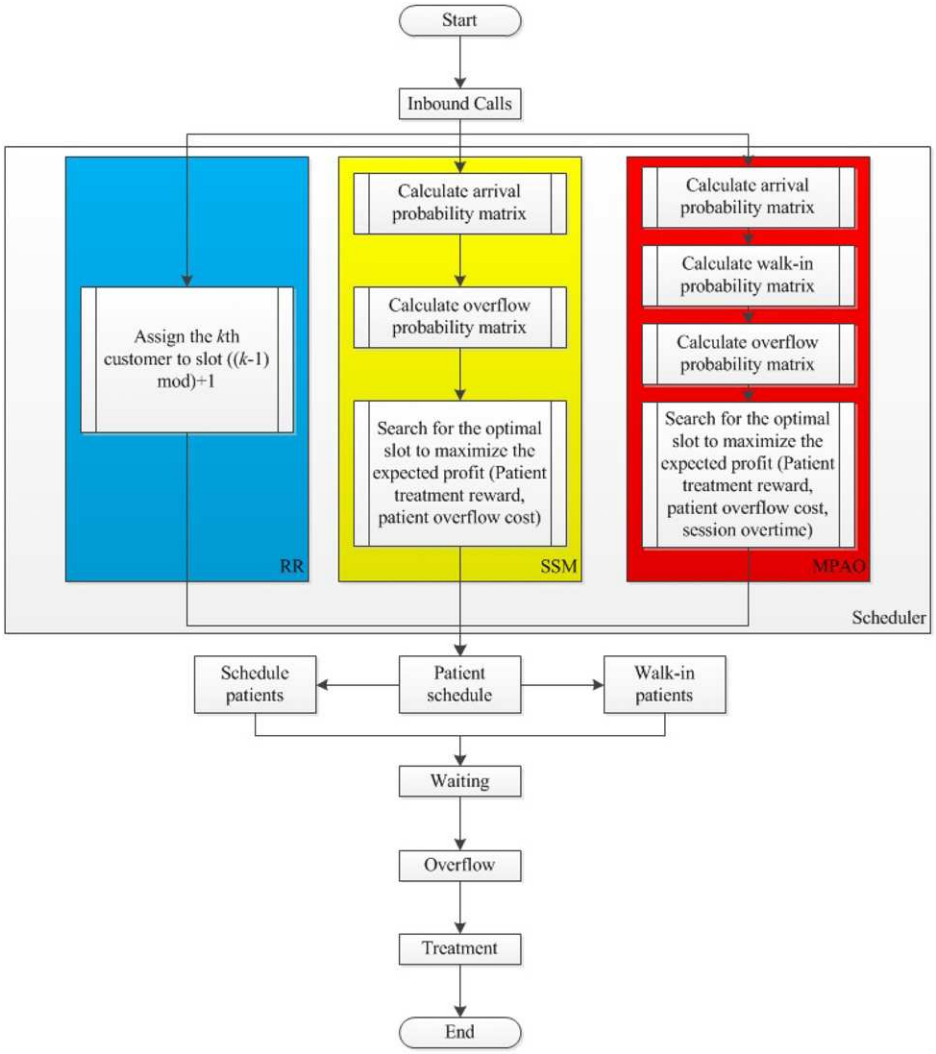

Fig. 5. Simulation-based experiment workflow.

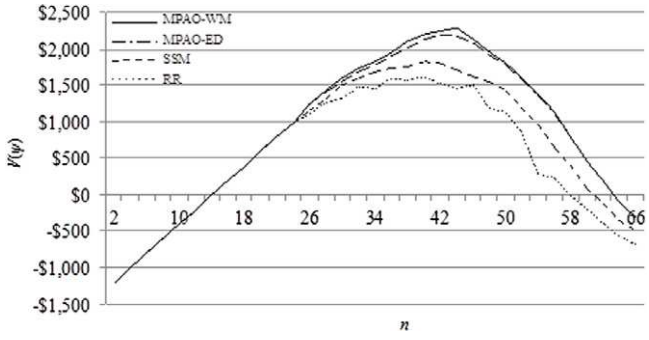

Fig. 6. Analysis of different outpatient scheduling methods ($I = 15, J = 2$ and $\alpha = 0.9$).

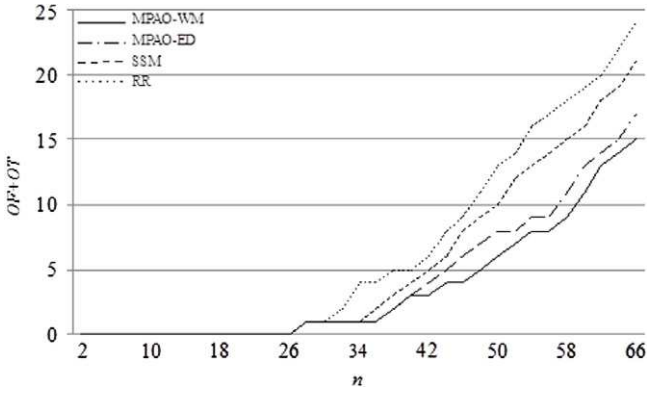

Fig. 7. Analysis of different outpatient scheduling methods in terms of the number of overflowing patients ($I = 15, J = 2$ and $\alpha = 0.9$).

all the models perform identically. However, when the number of scheduled patients increases and patients are overbooked, it is observed that, compared to the MPAO-ED and MPAO-WM methods, the probability of having overflowing patients for the RR and SSM methods increases.

These results indicate that the various scheduling methods become distinct when the expected profit function approaches the number of patients that generates the maximum expected profit. When additional patients are scheduled, resulting in an increase in patient overflow, the MPAO-WM models are found to be superior to the other scheduling approaches. The expected profits of different scheduling approaches are summarized in Table 6.

4.3. Analysis of clinic service time capacity

Typically, the number of appointments per day is determined based on the provider service time and the number of patients

Table 6
Summary of the expected profits.

Scheduling method	Maximum expected profit per provider ($)	Number of treated patients at max. profit
MPAO-WM	2290	44
MPAO-ED	2220	42
SSM	1820	40
Round-Robin	1590	36

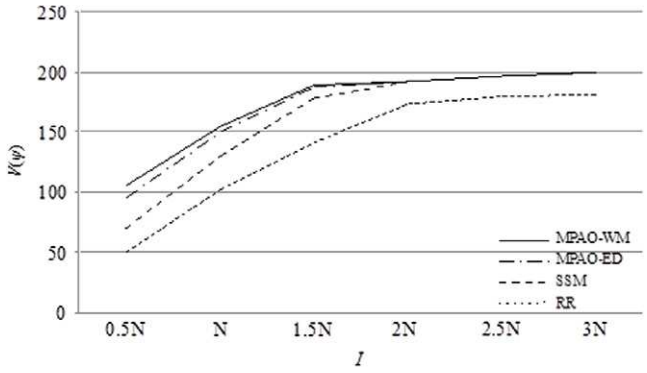

Fig. 8. Analysis of different outpatient scheduling methods ($N = 15, J = 2$ and $\alpha = 0.9$).

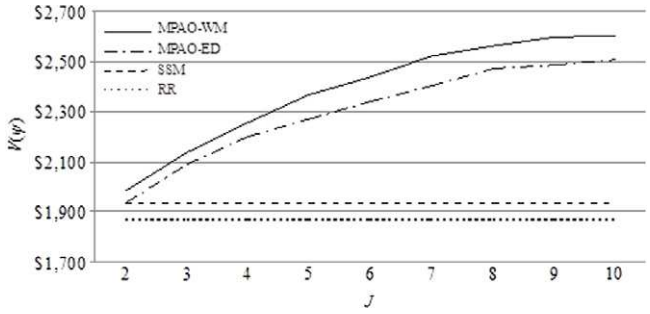

Fig. 9. Analysis of different outpatient scheduling methods when ($N = 15, J = 2$ and $\alpha = 0.9$).

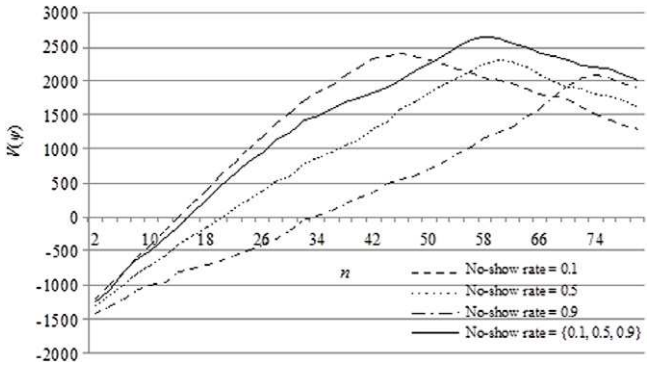

Fig. 10. Impact of no-show rates in MPAO-WM model when ($I = 15, J = 2$ and $\alpha = \{0.9, 0.5, 0.1\}$).

served in a day. A sensitivity analysis was performed to examine the impact of increasing I on the expected profit and the resulting behavior of the MPAO models. The analysis was performed by increasing $I = N = 15$ to $I = [0.5N, 3N]$ number of appointments as shown in Fig. 8. The results indicate that the expected profits of all the scheduling methods increase as I increases. Moreover, all scheduling approaches yield almost identical expected profit when $I > 2N$, which suggests that the expected cost of overflowing patients and overtime is close to zero due to a large capacity of the clinic. The results also indicate that increasing I or extending working hours has a limited impact on the expected profit when the patient demand is constant.

4.4. Analysis of multiple-provider capacity

To evaluate multiple-provider capacities, the MPAO-ED and MPAO-WM models have been analyzed with a varying number of providers (i.e., $J = [2, 10]$). It is assumed for each of the scheduling methods that all the providers have the same performance and treat the patients in the same manner. The results shown in Fig. 9 indicate that for both MPAO models the expected profit per provider increases as the number of providers increases, with the MPAO-WD yielding a better performance than the MPAO-ED in terms of the expected profit. On the other hand, the Round-Robin and SSM methods are not affected by increasing the number of providers, which can be attributed to the lack of patient redistributions. Furthermore, the results for over six providers indicate that expected profit is not affected significantly because the average expected profit approaches the maximum expected profit.

Based on the analyses of 10 providers, it has been found that the expected profit and the maximum number of scheduled patients per provider improve. Both expected profit functions from MPAO-ED and MPAO-WM behave similar with respect to the number of patients scheduled to reach the maximum expected profit with variation in the average expected profit per provider.

4.5. Analysis of patient no-show probability

It has been shown that the MPAO methods provide the maximum expected profit by scheduling a similar number of patients per provider, especially when the number of providers increases. To scrutinize this result further, the expected profit is analyzed when α, the probability of a patient attending an appointment, is varied in each experiment to the values: $\alpha = \{0.1, 0.5, 0.9\}$, which are uniformly distributed for all providers.

As shown in Fig. 10, the expected profit increases intuitively with the increase of $1 - \alpha$. The number of scheduled patients that reaches the maximum profit when α is fixed at $\alpha = 0.9$ is less than the expected profit curve when α is any lower value. Also, the least number of scheduled patients that still obtains the maximum profit occurs when $\alpha = 0.9$ for all patients. The maximum expected profit in all experiments is generated when α is selected to be a uniform random number from the set $\{0.1, 0.5, 0.9\}$. This result indicates that with an increase in the mixture of patient show probabilities, the scheduling system is able to place various patient types together to generate extra profit.

Although the maximum profit is obtained when $\alpha = \{0.1, 0.5, 0.9\}$, the scheduling system needs to assign more patients in the schedule when $\alpha = 0.9$. The results lead to the conclusion that a mixture of patient attendance probability does not affect the expected profit and actually has a positive effect on the number of patients served.

4.6. Analysis of walk-in patient arrival rates

The proposed algorithms can consider the no-show rates of scheduled patients and the arrival rates of walk-in patients

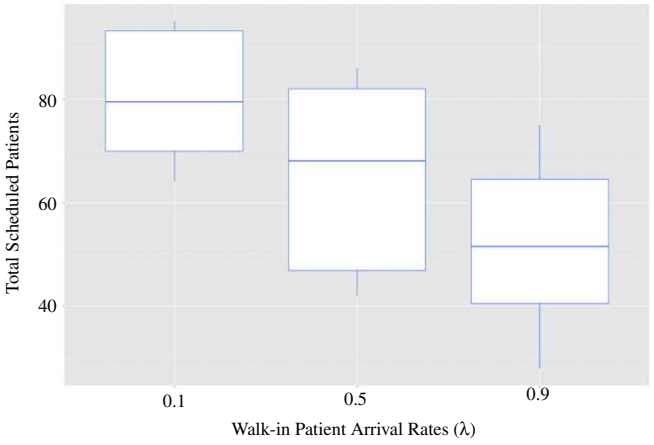

Fig. 11. Impact of walk-in patient arrival rate of MPAO-WM model on total scheduled patients when ($I = 16, J = 2$ and $\alpha = 0.1$).

into their solution procedures. In other words, to maximize the expected profit, the trade-off between no-shows and walk-in patients is captured by the proposed algorithms. As shown in Fig. 11, as the arrival rates of walk-in patients increase, the total number of scheduled patients is reduced. This result illustrates that the optimal schedule can balance the allocation slots between the no-shows and regular walk-in patients.

5. Conclusions and future work

In this research, an overbooking scheduling model for multi-provider outpatient clinics is proposed with the aim of maximizing their expected profit. A comprehensive objective function is proposed that considers a number of important measures in clinics, such as revenue from patients, penalties associated with patient waiting time, and staff overtime. The maximum expected profit function can be used as a natural stopping condition for the proposed scheduling algorithm. Furthermore, the scheduling algorithm's behavior was investigated with regard to the overbooking and load balancing methods. Simulation results indicated that the proposed model is superior to the RR and the SSM methods that are commonly utilized in outpatient clinics. Sensitivity analysis illustrated that the MPAO-WM model results in a greater expected profit than all other models, with respect to the varying clinic schedule capacity, number of providers, and patient no-show probability. The proposed model is applicable in multiple provider settings with fixed appointment duration and low patient-provider preference indicating that it is appropriate for implementation in dental hygiene services, medical rehabilitation services, and medical imaging services, such as magnetic resonance imaging (MRI), computed tomography (CT) scan imaging, and nuclear imaging. In the future, services that require variable service time such as endoscopy and cardiac catheterization, could be addressed by a modified MPAO model that considers variable service time and variations in the start and end time of procedures. The proposed model has not considered a patient preference for simplicity. To incorporate the considerations of care continuity, an extensive and comprehensive model to address patient preferences is encouraged to explore in future work.

References

[1] C. De Lathouwer, J.P. Poullier, How much ambulatory surgery in the world in 1996–1997 and trends? Ambulatory Surgery 8 (4) (2000) 191–210.
[2] K. Muthuraman, M. Lawley, A stochastic overbooking model for outpatient clinical scheduling with no-shows, IIE Trans. 40 (9) (2008) 820–837.
[3] T. Cayirli, E. Veral, Outpatient scheduling in health care: a review of literature, Prod. Oper. Manage. 12 (4) (2003) 519–549.
[4] D. Gupta, B. Denton, Appointment scheduling in health care: Challenges and opportunities, IIE Trans. 40 (9) (2008) 800–819.
[5] J.I. McGill, G.J. Van Ryzin, Revenue management: Research overview and prospects, Transp. Sci. 33 (2) (1999) 233–256.
[6] J. Patrick, M.L. Puterman, M. Queyranne, Dynamic multipriority patient scheduling for a diagnostic resource, Oper. Res. 56 (6) (2008) 1507–1525.
[7] N.T. Bailey, A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times, J. Roy. Statist. Soc. 14 (2) (1952) 185–199.
[8] J. Daggy, M. Lawley, D. Willis, D. Thayer, C. Suelzer, P.-C. DeLaurentis, A. Turkcan, S. Chakraborty, L. Sands, Using no-show modeling to improve clinic performance, Health Inform. J. 16 (4) (2010) 246–259.
[9] B. Zeng, A. Turkcan, J. Lin, M. Lawley, Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities, Ann. Oper. Res. 178 (1) (2010) 121–144.
[10] H. Balasubramanian, A. Muriel, L. Wang, The impact of provider flexibility and capacity allocation on the performance of primary care practices, Flex. Serv. Manuf. J. 24 (4) (2012) 422–447.
[11] A. Ozen, H. Balasubramanian, The impact of case mix on timely access to appointments in a primary care group practice, Health Care Manage. Sci. 16 (2) (2013) 101–118.
[12] B. Vanden, P. M, D.C. Dietz, Minimizing expected waiting in a medical appointment system, IIE Trans. 32 (9) (2000) 841–848.
[13] T.R. Rohleder, K.J. Klassen, Rolling horizon appointment scheduling: a simulation study, Health Care Manage. Sci. 5 (3) (2002) 201–209.
[14] R. Kopach, P.-C. DeLaurentis, M. Lawley, K. Muthuraman, L. Ozsen, R. Rardin, H. Wan, P. Intrevado, X. Qu, D. Willis, Effects of clinical characteristics on successful open access scheduling, Health Care Manage. Sci. 10 (2) (2007) 111–124.
[15] C.-J. Liao, C.D. Pegden, M. Rosenshine, Planning timely arrivals to a stochastic production or service system, IIE Trans. 25 (5) (1993) 63–73.
[16] S.A. Erdogan, B. Denton, Dynamic appointment scheduling of a stochastic server with uncertain demand, INFORMS J. Comput. 25 (1) (2013) 116–132.
[17] M. Murray, C. Tantau, Redefining open access to primary care, Manag. Care Quart. 7 (3) (1999) 45–55.
[18] M. Murray, C. Tantau, Same-day appointments: exploding the access paradigm, Fam. Pract. Manag. 7 (8) (2000) 45–50.
[19] F. Dexter, A. Macario, R.D. Traub, M. Hopwood, D.A. Lubarsky, An operating room scheduling strategy to maximize the use of operating room block time: computer simulation of patient scheduling and survey of patients' preferences for surgical waiting time, Anesth. Analg. 89 (1) (1999) 7–20.
[20] S. Kim, R.E. Giachetti, A stochastic mathematical appointment overbooking model for healthcare providers to improve profits, IEEE Trans. Syst. Man Cybern. A 36 (6) (2006) 1211–1219.
[21] S. Chakraborty, K. Muthuraman, M. Lawley, Sequential clinical scheduling with patient no-shows and general service time distributions, IIE Trans. 42 (5) (2010) 354–366.
[22] C.-J. Ho, H.-S. Lau, Minimizing total cost in scheduling outpatient appointments, Manage. Sci. 38 (12) (1992) 1750–1764.
[23] H.-S. Lau, A.H.-L. Lau, A fast procedure for computing the total system cost of an appointment schedule for medical and kindred facilities, IIE Trans. 32 (9) (2000) 833–839.
[24] E.J. Rising, R. Baron, B. Averill, A systems analysis of a university-health-service outpatient clinic, Oper. Res. 21 (5) (1973) 1030–1047.
[25] J.R. Swisher, S.H. Jacobson, J.B. Jun, O. Balci, Modeling and analyzing a physician clinic environment using discrete-event (visual) simulation, Comput. Oper. Res. 28 (2) (2001) 105–125.
[26] K.J. Klassen, T.R. Rohleder, Outpatient appointment scheduling with urgent clients in a dynamic, multi-period environment, Int. J. Serv. Ind. Manag. 15 (2) (2004) 167–186.
[27] T. Cayirli, K.K. Yang, S.A. Quek, A universal appointment rule in the presence of no-shows and walk-ins, Prod. Oper. Manage. 21 (4) (2012) 682–697.
[28] G.C. Kaandorp, G. Koole, Optimal outpatient appointment scheduling, Health Care Manage. Sci. 10 (3) (2007) 217–229.
[29] N. Liu, S. Ziya, V.G. Kulkarni, Dynamic scheduling of outpatient appointments under patient no-shows and cancellations, Manuf. Serv. Oper. Manag. 12 (2) (2010) 347–364.
[30] B. Sun, G.W. Evans, L. Bai, Simulation modeling and analysis of a multi-resource medical clinic, in: International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), IEEE, 2011, pp. 593–600.
[31] 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 J. Oper. Res. 183 (2) (2007) 812–826.
[32] R.R. Chen, L.W. Robinson, Sequencing and scheduling appointments with potential call-in patients, Prod. Oper. Manage. 23 (9) (2014) 1522–1538.
[33] T. Cayirli, E.D. Gunes, Outpatient appointment scheduling in presence of seasonal walk-ins, J. Oper. Res. Soc. 65 (4) (2013) 512–531.
[34] L.N. Cooper, D.S. Steinberg, Introduction to Methods of Optimization, W. B. Saunders Company, Philadelphia, PA, 1970.