Dynamic process modelling of patients' no-show rates and overbooking strategies in healthcare clinics

Article *in* International Journal of Engineering Management and Economics · January 2012

DOI: 10.1504/IJEME.2012.048602

2 authors:

Raid Al-Aomar
Abu Dhabi University

Mahmoud Awad
American University of Sharjah

Dynamic process modelling of patients' no-show rates and overbooking strategies in healthcare clinics

Raid Al-Aomar*

Engineering Management,
Abu Dhabi University,
P.O. Box 59911, Abu Dhabi, UAE
E-mail: raid.alaomar@adu.ac.ae
*Corresponding author

Mahmoud Awad

Industrial Engineering Department,
ALHOSN University,
P.O. Box 38772, Abu Dhabi, UAE
E-mail: m.awad@alhosnu.ae

Abstract: Clinic appointment cancellation and 'no-shows' by patients can disturb day-to-day scheduling and lead to inefficient resource allocation and lost revenue. This paper presents an approach for dynamic process modelling and analysis of patients' no-show rates in conjunction with the overbooking strategy. To this end, a dynamic discrete event simulation model of a clinic operation is developed to assess current clinic performance under various no-show rates and patterns and test several overbooking strategies to improve clinic's efficiency and enhance performance. The approach significance includes the analysis of variability in scheduling and treatment times, the impact of patients reneging, and walk-in patients. In addition, the paper develops a unified clinic utility function that analyses the impact of overbooking on clinic revenue (patients treated), capacity (utilisation), and customer service (wait time). An example of a general practice (GP) healthcare clinic is used to illustrate the development and application of the proposed approach.

Keywords: process modelling; appointment scheduling; healthcare clinics; no-show rates; overbooking; simulation; engineering management.

Reference to this paper should be made as follows: Al-Aomar, R. and Awad, M. (2012) 'Dynamic process modelling of patients' no-show rates and overbooking strategies in healthcare clinics', *Int. J. Engineering Management and Economics*, Vol. 3, Nos. 1/2, pp.3–21.

Biographical notes: Raid Al-Aomar is an Associate Professor of Industrial Engineering and the Director of Master of Engineering (MEM) programme at Abu Dhabi University. He holds a PhD in Industrial Engineering/Operations Research from Wayne State University. He has over 15 years of experience in the field of industrial engineering at companies and universities in Jordan, USA, and UAE with over 30 journal and conference papers and several book publications. His research interests include simulation-based optimisation and lean Six Sigma systems.

Mahmoud Awad received his BS in Mechanical Engineering from Jordan University of Science and Technology, Irbid, Jordan and MS in Industrial Engineering from University of Jordan, Amman, Jordan. He earned his PhD in Industrial Engineering from Wayne State University, USA. He worked as a Reliability Engineer at Ford Motor Company and as a Master Black Belt at Case New Holland. Currently, he is an Assistant Professor of Industrial Engineering at ALHOSN University, UAE. His research interests include robust design, experimental design, Six Sigma, and the application of quality improvement in healthcare systems.

This paper is a revised and expanded version of a paper entitled 'Dynamic process modeling of patients' no-show rates and overbooking strategies in healthcare clinics' presented at the Second International Conference on Engineering Systems Management and Applications (ICESMA 2010), American University of Sharjah, UAE, 30 March to 1 April 2010.

1 Introduction

The healthcare industry has been going under intense pressure recently to change its current practices and undergo major reforms due to its contribution to national economic growth and the ever increasing demand for better patient quality service. Currently, the USA's cost of healthcare is estimated to be 15% of the GDP and will be exceeding 20% by 2015. Among the reasons for the high cost is raw material inventory, capacity planning, and customers' no-show or cancellation rates.

As most real world processes, healthcare systems are characterised by uncertainty, variability, dynamic and complex behaviour. Dynamic process modelling (DPM) is, therefore, commonly used for process management and business process reengineering of healthcare systems including hospitals, clinics, scheduling, and other related operations. Process modelling involves the application of modelling tools (analytical and graphical) to acquire process knowledge and represent process structure and operation. Commonly used process models provide a graphical description of the process elements, activities, resources, and relationships using a process diagram (i.e., a process map or flowchart). As discussed in Jacobson et al. (2006), the objective of process modelling is to better understand the process and to provide a mean for process design and improvement.

Simulation modelling in general and discrete event simulation (DES) in particular has played an increasing role in healthcare process improvement, design, and problem-solving. This can be mainly attributed to the tremendous improvement in simulation software and the integration of statistical and optimisation methods (Banks, 1998; Kosfeld, 1998; Law, 2006; Nelson et al., 2006). Many applications of process modelling in healthcare industry are reported in the literature (Boxerman, 1996; Colli et al., 2007; Jun et al., 1999 Vemeulen et al., 2009). In an effort to maximise the patient throughput and reduce patient waiting time in the system, Ahmed and Alkhamis (2009) integrated simulation with optimisation to design a decision support tool for the operation of an emergency department unit at a governmental hospital resulting in significant improvements.

The no-show rate (i.e., patients failing to appear for scheduled appointments) and cancellation issue have been under scrutiny by many researchers lately due to its significant impact on cost and resource allocation. Moore and Wilson-Witherspoon

(2001) reported that health providers' loss is around 3% to 14% of total clinic income due to patients' non-attendance in the USA. Thus, several overbooking strategies have been proposed to stabilise revenues and reduce costs related to no-show rates. These strategies, however, raise the concern of increasing waiting time and reneging which triggers patient dissatisfaction (Harper and Gamlin, 2003). Campbell et al. (2000) reported that 94% of the variance in patient's dissatisfaction with practices that resulted in appointment waiting times in excess of 15 minutes. A recent consumer report survey ranked patient waiting time as the number one complaint among patients (Consumer Reports, 2007). Another important concern is that bad waiting experiences trigger anger toward the clinic and can be a major reason for future no-show ups (Zeng et al., 2009).

This paper adopts a dynamic DES model of a clinic operation to assess the clinic performance under various no-show rates and to test overbooking strategies. The proposed objective is to improve clinic efficiency at a reasonable level of patient satisfaction. Section 2 reviews the literature on no-show rates and overbooking strategies, Section 3 states the problem, Section 4 presents the solution approach, Section 5 discusses a case study, and Section 6 presents the study conclusions.

2 Literature review

Studies tackling the no-show issue can be categorised into three main categories:

1. Research focused on estimation of no-show rates and studying factors impacting no-show rates (Moore and Wilson-Witherspoon, 2001; Deyo and Inui, 1980; Dove and Schnieder, 1981; Satianti et al., 2009).
2. Review of clinics historical records to quantify the financial impact of no-show (Goldman et al., 1982; LaGanga and Lawrence, 2007; Moore and Wilson-Witherspoon, 2001).
3. Providing potential solutions using different strategies or numerical scheduling solutions (Bech, 2005; Brahimi and Worthington, 1991; Jain and Chou, 2000).

An initial study suggested that patient demographics, medical conditions, and environment are major indicators of no-show behaviour (Deyo and Inui, 1980). Woodcock (2003) reported the no-show issue, as one of the most serious operational issues, with a rate, ranging from 5% to 60%. A more recent study reported a no-show and/or cancellation rate in the range of 8% to 15% for a dental clinic (Abdus-Salaam and Davis, 2010). Reported reasons for no-shows include lack of transportation, scheduling problems, oversleeping or forgetfulness, and lack of child care (Cayirli and Veral, 2006). In another study, Penneys and Glaser (1999) studied the cancellation and no-show rates in a dermatology clinic over a six months period for 4,876 patients. They concluded that same day cancellation was 8.3% and no-show rate was 17% with no difference due to gender or visit type (first or follow up). The problem extends to medical labs as well (Moore and Wilson-Witherspoon, 2001; Lacy et al., 2004).

Although there has been a lot of research on no-show and cancellation rates, very few researchers provided practical solutions to the problem. According to Bech (2005), health management has tackled the cancellation and no-show up problem using several customer-based solutions such as: imposing fines, providing incentives, overbooking, reminder systems, and customer punishment (being discharged from the waiting list). However, he argued that very few studies include more than one intervention and very few report information on the cost of the intervention which enables only vague conclusions. He reported two studies where no-show rate has decreased from 6.4% to 5.5% after the introduction of a fine to non-attendance and from 20.1% to 9.27% after introducing $30 fine fee. He argued that the observation period in both studies is short which may differ from a long term effect of introducing a fee. For the same objective, Satianti et al. (2009) monitored the no-show rate nine months before an automated telephone reminder was put in place and 17 months after and reported no significant reduction. Similarly, Kheirkhah et al. (2009) studied the no-show problem and the effect of implementing a reminder system in more than 10 VA hospitals and noticed no significant effects on reducing no-show rates. Jain and Chou (2000) claimed that a new patient hour-long orientation to clinic decreased the new patient no-show rate significantly from 45% to 18%. In another study, Hamilton et al. (1999) found that providing patients with an information pamphlet or a copy of their referral letter did not affect no-show rates.

2.1 Patients overbooking strategies

Under the lack of an effective strategy to handle the financial impact of the no-show, researchers advocated different scheduling schemes such as static overbooking (Kim and Giachetti, 2006; Denton and Gupta, 2003; Greasley, 2006; LaGanga and Lawrence, 2007; Robinson and Chen, 2003), open access scheduling (Kennedy and Hsu, 2003) and dynamic overbooking (Muthurman and Lawley, 2008; Chakraborty et al., 2011). Cayirli and Veral (2006), and Gupta and Denton (2008) provided an extensive review of scheduling schemes. The majority of these schemes consider a multi-objective optimisation model that balances between revenue and overtime and the cost of sending patients away. These strategies or scheduling schemes can be divided into two groups: static and dynamic. In static, the number of appointments are known ahead of the visitation session, where in dynamic, schedules are evolving depending on appointment calls. As an example of static models, Kim and Giachetti (2006) proposed a model with a total expected profit per physician, M, given number of appointments $A = a$ as follow:

$$\begin{aligned} &E\left(M \mid A=a\right) = \text{(revenue generated)} - \text{(cost incurred)} \\ &= \sum_{w}\sum_{s}\left\{R\min(s+w, U) - Y - C_o \max\left[\min(s+u, U) - N, 0\right]\right. \\ &\left. -C_e \max(s+w-U, 0)\right\} \times P\left(S=s \mid A=a\right) P(W=w). \end{aligned} \tag{1}$$

where

- r average revenue generated per patient.
- m discrete random variable representing number of patients, who show up for their appointments.
- W discrete random variable representing the number of patients who walk in without an appointment.
- U the upper limit number of patients to see on regular time plus overtime such that ($U \geq N$).

N number of appointments during regular time.

Y total fixed cost per physician including facility, salaries of physician nurse and administration during regular time regardless of number of patients.

C_o overtime cost per patient.

C_e penalty cost per patient incurred when number of patients showing up is beyond the upper limit capacity, i.e., $m + W > U$. C_e represents the cost of sending a patient away.

The model in equation (1) assumes constant no-show and walk-in probabilities. It also assumes that the total revenue increases proportionally as $m + w$ increases up to U patients ($r\min(m + w, U)$. If the number of patients exceeds U, no more revenue is generated. In addition over time cost is incurred ($C_o\max[\min(m + w, U) - N, 0]$) only when $N < m + w < U$. When $m + w > U$, a penalty cost of sending patients away is incurred ($C_e \max(m + w - U, 0)$).

The above model, as the case in all static scheduling schemes, suffers from few shortcomings. For example, the model assumes N number of appointments ahead of time, with a single probability of no-show for all patients, and appointments once scheduled are not adjusted during the day. To overcome these shortcomings, open access and dynamic scheduling schemes are proposed. In open access scheduling, patients are allowed to get appointments within a day or two of when they make the appointment call. In the dynamic case, patients call in for appointments, are screened, assigned a probability of no-show, and finally, given an appointment using an objective function that captures patient revenue, staff overtime cost, and patient waiting time. As an example of dynamic scheduling, Muthurman and Lawely (2008) proposed dividing a single service period (typically a day) into I intervals each called a 'slot' where each slot $i = 1, 2, \ldots, I$ is of length Δt_i. Patients, on the other hand, call in to schedule an appointment before the beginning of slot 1. When a patient calls in, a probability j will be determined based on the patient attributes and clinic's no-show model, and then the optimum i^* slot is chosen as follow:

$$i^* = \arg\max f(Q, R),$$

where

$$f(Q, R) = r\sum_i \sum_m mQ_{i.m} - \sum_i c_i \sum_k kR_{i.k} \tag{2}$$

where

Q expected number of patients showing up for their appointments.

R expected number of patients who show up for their appointments but could not be served within their appointments and have *t overflow* to the proceeding slots to be served.

r average revenue generated per patient.

m discrete random variable representing the number of patients, who show up for their appointments.

$Q_{i,m}$ probability of m patients arriving at slot i.

C_i penalty costs for not serving a patient within his appointment slot i. Usually C_i is higher if patient is served outside the regular working hours.

K discrete random variable representing the number of patients, who show up for their appointments and overflow to the proceeding slots to be served.

$R_{i,k}$ probability of k patient's overflow from slot $i - 1$ to i.

The model assumes a punctual arrival time and exponential inter-arrival time with no walk-ins. Although the number of overflows to next slot (i.e., R is a good surrogate to waiting time), it is not exactly the same. This will be discussed in Section 5.2. Chackraborty et al. (2010) used the same model but with Gamma service arrival time based on the fact that physicians can adjust the service time based on patients' log.

LaGanga and Lawrence (2007) evaluated the utility of appointment overbooking as a means to mitigate the negative impact of patient no-shows in healthcare clinics, where no shows reduce provider productivity and clinic efficiency, increase healthcare costs, and limit patient access to necessary services. They presented a utility model to evaluate appointment overbooking in terms of the trade-offs between the benefits of serving additional patients and the costs of increased patient wait time and provider overtime. It is based on adjusting subjective weights for the relative importance of patient access, patient wait times, and clinic overtime. Results indicate that patient overbooking can often lead to substantially improved net clinic performance. Other examples of scheduling policies include double-booking (Rohleder and Klassen, 2006) and wave scheduling (Chung, 2002). In double-booking, a clinic schedules two appointments to start at the same time and in wave-scheduling, uses cyclical booking patterns to improve clinic performance.

3 Problem statement

The increasing healthcare cost and the growing pressure on healthcare providers to maintain high quality of the delivered medical services have forced many clinics to consider various structural and operational changes. This includes the need to address the issue of increasing no-show up rates and the application of systematic approaches that result in better utilisation of clinic resources and increased revenues while maintaining an acceptable level of provided services. As discussed in the literature review, clinics have frequently adopted overbooking as a strategy to deal with their prevalent financially-related patient no-show and cancellation problem. However, past research did not address the impact of the variability of scheduling and treatment time, patients' reneging, and walk-in policies on the effectiveness of the overbooking strategies. The impact of not addressing these matters may result in underutilised resources/lost revenues, or lost customers due to long waiting time.

In terms of expected number of patients showing up (Q), expected waiting time (Wt), and reneging (RE), the problem addressed in this research can be stated as follows:

$$Max\ f(Q, Wt, RE) = r\sum_{i=1}\sum_{m} mQ_{i.m} - \sum_{i=1} C_{wi}\sum_{m} Q_{i,m}Wt_{i.m} - C_{ren}\sum_{i=1}\sum_{m} Q_{i,m}RN_{i.m} \quad (3)$$

where

Wt expected waiting time for all patients showing up for their appointments.

RE expected number of patients who will abandon the waiting queue due to long waiting time (Wt > 45 minutes).

C_{wi} penalty costs for patient waiting for his/her appointment per minute in slot i. Usually is higher if patient waited outside the regular working hours.

$Wt_{i.m}$ total waiting time for all patients scheduled for slot i when m patients show up.

C_{ren} cost of loosing patients due to number of reneged patients (RN).

$RN_{i.m}$ number of reneged patients due to long waiting time (Wt_i < 45 minutes) scheduled for slot i when m patients show up.

The 45 minute reneging rule was selected based on a quick patient survey where patients with appointments were asked regarding the longest waiting time they can tolerate before leaving the clinic.

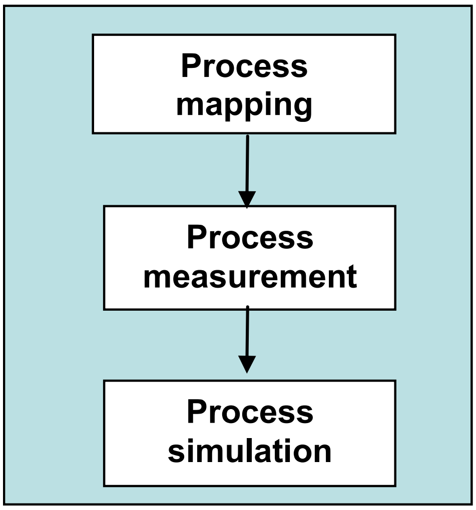
Solving the above model analytically is challenging due to several reasons, among those reasons are the stochastic nature of some of the variables such as the randomness nature of inter-arrival time and service time in addition to the probability of no-show, reneging, and walk-in. Hence, DES will be used to investigate different overbooking strategies and study the effectiveness of each one using several performance measures; namely waiting time, throughput, and utilisation.

4 Solution approach

As shown in Figure 1, DPM approach consists of three main steps; process mapping, process measurement, and process simulation. In process mapping, we conceptualise and characterise the underlying process in terms of content and logic where a process flow chart or block diagram is used to map the process. In process measurement, we collect pertinent data and measure the process in terms of time, labour, and cost. The process map (structure) and measurements (data) are then used to develop a process simulation model. Process simulation is the platform for assessing process performance measures and testing improvement actions. The approach is focused on gaining process knowledge and improving clinic operations. In this paper, however, the focus will be on studying the impact of various no-show rates and measuring the resulting improvement from using overbooking strategies at various rates.

The proposed approach can be described by a general DPM model (shown in Figure 1) and a detailed clinic model (shown in Figure 2). The approach is set to be simulation-based and to handle the issues outlined in the problem statement.

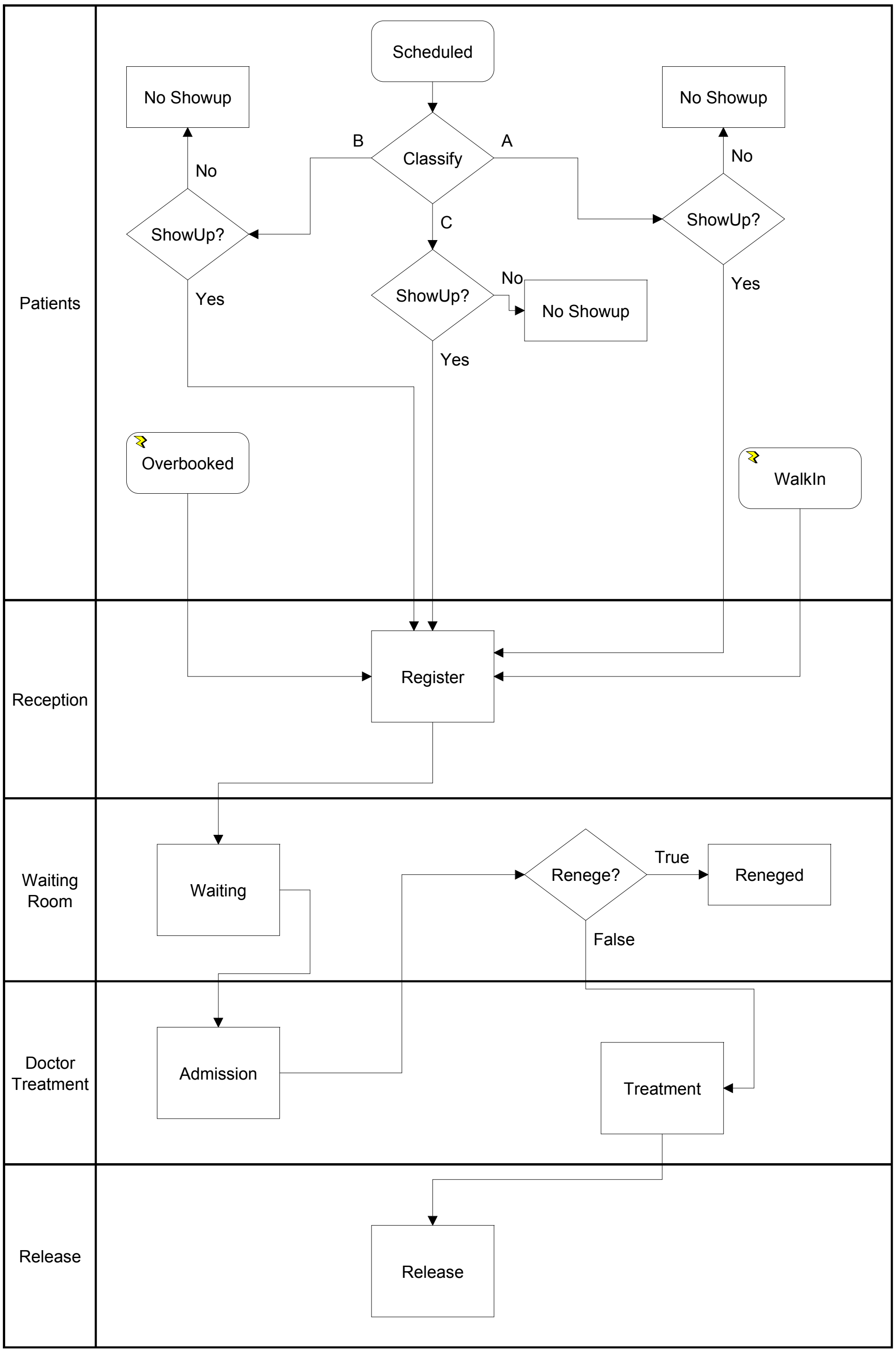
Figure 1 DPM approach (see online version for colours)



The proposed approach emphasises the role of DES in dynamic modelling and analysis of the overbooking strategy selection for healthcare clinics. The simulation model of the general practice (GP) healthcare clinic is developed using iGrafx™ (2006) process modelling and simulation software. The mapping, modelling, and simulation features of iGrafx™ allow for drawing a process diagram, creating a process model, simulating the process, and then reviewing the process results in a simulation report. It provides an excellent environment to model process details (cycle times, resources, operating pattern, and costs) and run dynamic simulations of the process flow. iGrafx™ allows modelling activity times using probability distributions. A number of test runs were undertaken to verify the model. In addition, the software has a built-in module for design of experiments (DOE) which allows for running simulation-based experiments at different settings of overbooking strategy. Figure 2 shows the developed simulation model of the GP clinic.

The proposed model in Figure 2 consists of four flexible patient modules; scheduling, classification, overbooking, and reneging. The scheduling module generates a full-day patient arrival schedule (e.g., 32 slots of 15 minute each for 8 hours). To model the actual behaviour of the patients, their inter-arrival times follow an exponential distribution with the same mean. Decision points are used to model show/no-show probability and walk-in accept/reject policy. The classification module is used in case patients' arrival is dependent on certain predetermined no-show probabilities. The overbooking module is used to incorporate the overbooking strategies used in patient scheduling. The reneging module is finally used at the patient waiting stage to model patient behaviour in leaving without treatment when the waiting time is long. Attributes from these modules follow the patients flow from registration to release.

Figure 2 Process simulation model of GP clinic (see online version for colours)



The following notation is used in the proposed approach:

WI walk-in

NS no show-up

OB overbooked

RN reneged

DT daily treatments

WT wait time (min)

DU doctor utilisation.

In the proposed approach, the effectiveness of overbooking, along with other strategies, are investigated at a GP health clinic with the objective of improving overall clinic performance in terms of the following:

1 number of daily treated patients

2 patient waiting time

3 the utilisation of medical staff (doctors and nurses).

Equation (3) includes the cost elements for waiting, overtime (if applicable), and reneging as a lost customer might not come back. In addition, other approaches account for clinic overtime when patients are served outside the regular timing of the session. In the proposed approach no cost numbers are used and the overtime and reneging and incorporated dynamically into the model. The waiting time will indicate the number of reneged patients (RN) and the number of daily treated patients indicates the number of those who leave without treatment at the end of regular hours (end of simulation run). To estimate the number of reneged patients, we can calculate the probability $p(Wt > 45 \text{ min})$.Using an exponential distribution of service time with mean service time of 15 minutes, p is about 0.05 resulting in a service level of 95%. Therefore, out of 32 scheduled per day who arrived on time, it is expected that 0.8 patient on average will leave the clinic each day without treatment.

Simulation results will show the number of reneged patients and the patients' overflow. Estimating cost values for reneging, waiting, and overtime is not sufficient for planning overbooking strategies. The focus of our approach is on clinic's overall utility in terms of productivity, utilisation, and service level. To this end, a unified utility function is developed by combining multiple criteria in a normalised score.

The utility function is a weighted average of revenue (in terms of DT), capacity (in terms of DU), and service level (in terms of WT). The service level is set based on customer surveys which indicated that patients will be satisfied if their total time in the clinic did not exceed 1 hour (A maximum of 45 min waiting and 15 min treatment). After 45 minutes of waiting, patients are assumed to leave the clinic without treatment (reneging). Using equal weights and normalisation, the three KPIs are combined into an overall score that ranges from 0 to 100. This provides a criterion for optimising clinic scheduling and overbooking that is better than assumed revenues and the costing of customer wait time, overflow, and over time. The utility formula can be customised to suit clinical operation and service level. In our case study, the formula of normalised utility is as follows:

$$U = \frac{1}{3}\left(\frac{DT}{32}\right) + \frac{1}{3}(DU) + \frac{1}{3}\left(\frac{45 - WT}{45}\right) \quad (4)$$

The simulation model is set to report clinic performance in terms of three KPIs along with the normalised utility score *U*. In addition, the model reports WI, OB, NS, and RN. The approach application can be set in two stages. First, a current state clinic analysis is conducted by assessing current schedule and analysing the impact of variability, reneging, and no-show. Second, a future state clinic operational model is proposed in which walk-in and overbooking strategies are utilised to improve the overall clinic utility.

5 Case study results

The DPM approach is applied to a GP healthcare clinic example. Data collected from a US-based GP clinic include: operating from 8 AM to 5 PM with one-hour lunch from 12 to 1 PM, scheduling patients every 15 minutes, and treating a patient every 15 minutes on average. Three levels of no-show up rates are used (low: 5%, medium: 15%, and high: 30%). Patients' reneging is also modelled where some patients may enter clinic and leave before treatment due to long waiting time. The primary focus is on measuring dual responses; physicians' utilisation and patients' waiting time in addition to the number of treated patients. Several scenarios were simulated using the clinic model, those scenarios are:

1. Baseline/ideal: measuring the performance of clinic with one (GP) physician with no overbooking strategy or any other no-show up fix strategy.
2. No-show up: the impact of various no-show-up rates on clinic performance.
3. Walk-in: testing the impact of accepting walk-in patients on clinic performance.
4. Overbooking: testing the impact of applying various rates of patients' overbooking on clinic performance.

The process simulation model of the example GP clinic is shown in Figure 2. The model shows the blocks (flow chart) at various stages/services of the GP clinic. Model data (time, resources, cost, etc.) is inserted by detailing each block. Nine scenarios are modelled to analyse the current clinic performance. In this terminating simulation (i.e., the model is reset at the end of working day), a 9-hour run time is used with multiple runs (five replications).

5.1 Current state clinic analysis

The simulation model is first used to assess the current state of clinic operations. To this end, eight scenarios are simulated. At each scenario, the values of the three KPIs (DT, WT, U) and the corresponding clinic utility score *U* are reported in addition to the patients' counts at four categories (WI, RN, NS, OB). Starting from an ideal scenario, the analysis will address several operational issues including variability, reneging, and no-show. Common no-show rates are first analysed (5%, 15%, and 30%). Table 1 summarises the simulation results (clinic performance measures) at the 9 scenarios along with the assumptions used in each scenario.

As shown in Table 1, the ideal scenario of clinic operation which results in treating almost all scheduled patients per day (31 out of 32). One patient remained in the clinic at day end due to simulation nature. Practically, treatment continues and the last patient is released. Waiting time is negligible and the doctor is fully utilised. When treatment time is subject to variability, utility drops due to 75% compared to 67% at variability in patients' inter-arrival time. This highlights the importance of service consistency. When variability in both arriving and treatment times is incorporated, the clinic operations are disturbed resulting in treating only 26 patients per day on average with longer waiting time and reduced utilisation (around 70%). With a full-show of patients, adding the element of patient reneging (i.e., patients leave without treatment when waiting time exceeds 45 minutes) to variability and has little impact on the number of treatments per day while reducing the waiting time and the doctor utilisation.

Table 1 Simulation results of 8 clinic scenarios (diagnosis of current state)

Scenario		*Assumptions*	*Patients*				*Performance measures*			*Utility*
			WI	*RN*	*NS*	*OB*	*DT*	*WT*	*DU*	*U*
1	Ideal	Scheduled patients arrive and treated on time	0	0	0	0	31.0	1.9	98.9%	97.18
2	Service variability	Exponential service times	0	0	0	0	27.0	19.6	84.3%	75.04
3	Arrivals variability	Exponential inter-arrival times	0	0	0	0	28.0	32.9	88.9%	67.76
4	Variability impact	Exponential inter-arrival and service times	0	0	0	0	26.0	23.5	80.3%	69.78
5	Reneging impact	Patients time leave without treatment	0	1	0	0	26.0	18.2	78.9%	73.24
6	With 5% no-show up	5% of scheduled patients will not show-up	0	3	2	0	24.0	15.0	70.7%	70.79
7	With 15% no-show up	15% of scheduled patients will not show-up	0	2	5	0	23.0	13.6	64.5%	68.72
8	With 30% no-show up	30% of scheduled patients will not show-up	0	1	10	0	18.0	9.9	54.7%	62.98

Scenarios 5 through 8 analyse the impact of no-show under variability and reneging conditions. Three levels of no-show rates are tested (5%, 15%, and 30%). It is first notices that as the percentage of patients' no-show increases the number of reneged patients decreases and customer service in terms of waiting time improves. However, fewer patients will be served with less utilisation of resources resulting in low clinic utility. At 30% no-show rate (i.e., 10 patients will not show up), the clinic treats only 18 patients per day resulting in only 55% utilisation and 63% overall utility. This clearly highlights the impact of no-show-up on clinic performance especially at high no-show rates. Such low clinic performance calls for testing and deployment of remedial actions.

5.2 Future state clinic analysis (remedial actions)

Once the diagnosis stage is completed, remedial actions are planned to face Table 1 scenarios which are likely to affect clinic operations. This includes relying on walk-in patients to compensate for no-show-ups, naive overbooking (i.e., overbooking the same rate of no-show up), and combining walk-in with overbooking. Accepting walk-in patients and setting overbooking strategies are both aimed at improving clinic performance and achieving a better future state. A walk-in patient is expected every 2 hours on average (on average three walk-in patients per day). Some clinics will have to accept walk-in patients as emergency cases. In such case, relying on walk-in will not be considered a remedial strategy and will be included as a source of variability in the model.

Table 2 summarises the results of remedial strategies. Each remedial strategy is tested at the three no-show-up rates resulting in nine simulations.

Table 2 Simulation results at walk-in and overbooking strategies

Simulated remedy		*Patients*				*Performance measures*			*Utility*
		WI	*RN*	*NS*	*OB*	*DT*	*WT*	*DU*	*U*
1	Accept walk-in patients at 5% no-show-up	3	3	2	0	26.0	23.6	76.6%	68.47
2	Accept walk-in patients at 15% no-show-up	3	2	5	0	24.0	15.5	70.7%	70.42
3	Accept walk-in patients at 30% no-show-up	3	1	10	0	23.0	12.9	64.5%	69.24
4	Naive: 5% overbooking at 5% no-show-up	0	1	2	2	26.0	21.9	76.6%	69.73
5	Naive: 15% overbooking at 15% no-show-up	0	2	5	5	26.0	20.8	76.6%	70.54
6	Naive: 30% overbooking at 30% no-show-up	0	1	10	10	26.0	22.5	78.8%	70.02
7	Naive: 5% overbooking + walk-in at 5% no-show up	3	4	2	2	26.0	23.6	76.6%	69.09
8	Naive: 15% overbooking + walk-in at 15% no-show-up	3	5	5	5	26.0	22.7	76.6%	69.14
9	Naive: 30% overbooking + walk-in at 30% no-show-up	3	4	10	10	26.0	22.7	78.8%	69.87

As shown in Table 2, accepting walk-in patients has a marginal impact at low no-show rate. However, at high now-show rate (30%), accepting walk-in patients has increased treated patients from 18 to 23 and utilisation from 55% to 65% and consequently improved the overall utility from 63% to 69%. However, this is not sufficient due to low utilisation. Overbooking strategy is therefore used to compensate for no-show patients and improve utilisation. Naive overbooking is a little better than relying on walk-in patients. It has improved the overall utilisation to 70% under the same conditions. Compared with baseline, walk-in solution is not effective for 5% no-show-up rate. Similarly, overbooking in the case of 5% no-show is not justified.

It is also noticed that the majority of remedial strategies had resulted in an overall utility of about 70% (26 patients treated daily, around 22 minute average waiting time, and around 76% utilisation). Reduced utilisation is compensated by improved service level (reduced waiting time). A key question to answer; is treating 26 patients daily along with about 75% to 80% utilisation good enough? If so, the 70% overall utility may be accepted. Otherwise, further improvement actions are needed. The clear opportunity is the reneged patients. If the 4 or 5 reneged patients are motivated to stay for treatment, the utilisation and daily treated patients will improve pushing the overall utility to about 80%. Another opportunity is reducing variability in service and inter-arrival times. Table 3 summarises the achieved improvement using the three remedial actions at the three now-show rates.

Table 3 Summary of improvement achieved at remedial strategies

	Measures	*Current state*	*Walk-in (WI)*	*Overbooking (OB)*	*OB + WI*
5% no-show	DT	24.0	26.0	26.0	26.0
	WT	15.0	23.6	221.9	23.6
	DU	70.7%	76.6%	76.6%	76.6%
	Overall utility U	*70.8*	*68.47*	*69.73*	*69.09*
15% no-show	DT	23.0	24.0	26.0	26.0
	WT	13.5	15.5	20.8	22.7
	DU	64.5%	70.7%	76.6%	76.6%
	Overall utility U	*68.98*	*70.42*	*70.54*	*69.14*
30% no-show	DT	18.0	23.0	26.0	26.0
	WT	9.9	12.9	22.5	22.7
	DU	54.7%	64.5%	78.8%	78.8%
	Overall utility U	*62.98*	*69.24*	*70.02*	*69.87*

5.3 Dynamic scheduling with patient-dependent no-show rates

In this section, we analyse an overbooking strategy in which overbooked patients are distributed at scheduled time slots depending on the type/class of the scheduled patient at the corresponding slot. In their study, Muthuraman and Lawely (2008) demonstrated their proposed stochastic method using an example where a visitation session is divided into eight slots ($i = 1 - 8$) with 1/2 hour for each slot. Patients are also divided into three groups ($j = 1 - 3$) with probabilities of show up of 0.25, 0.5, and 0.75, respectively. The final schedule for 34 patients' calls is summarised in Table 4.

We conducted a simulation of the above schedule, where the state of patients (i.e., show or no-show) is simulated using a binomial number with a probability of success equivalent to the probability of show up), and service time as an exponential distribution with λ = 10 minutes per patient. The simulation was conducted five times; Figure 3 summarises the results of the waiting time for each patient that showed up to his/her appointment on time:

Table 4 Dynamic scheduling scenario

Appointment slots i	*Patients groups j*		
	j = 1	*j* = 2	*j* = 3
1	2	2	2
2	2	1	2
3	1	1	2
4	2	2	1
5	0	1	2
6	4	1	1
7	0	0	3
8	1	0	1

Source: Muthuraman and Lawely (2008)

Figure 3 Dynamic scheduling simulation results

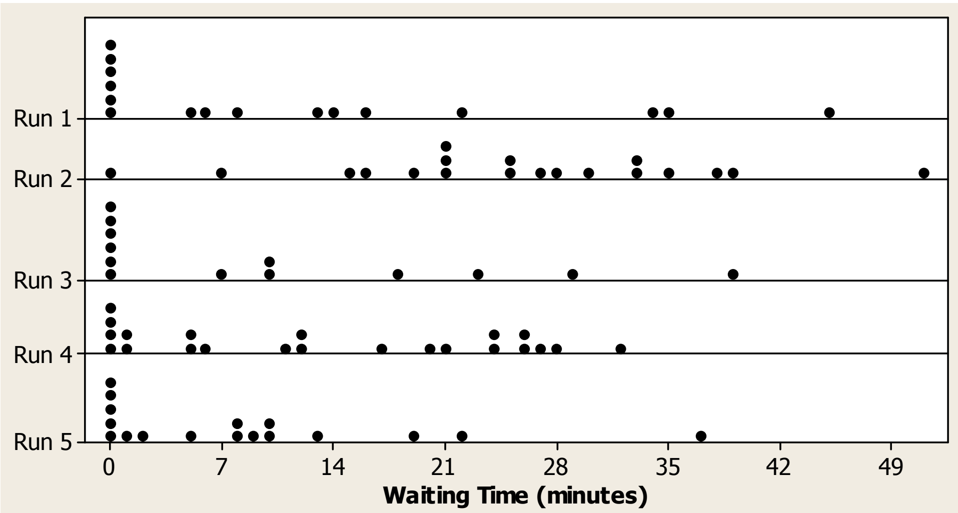


Table 5 Summary of dynamic scheduling performance

Measures	*DT*	*WT*	*WT-overtime*	*DU*	*RN*
Run 1	16	12.4	0	66.7%	1
Run 2	19	25.5	0	79%	1
Run 3	13	10.3	0	54%	0
Run 4	22	13.5	1	91%	0
Run 5	17	8.4	0	70%	0

As shown in Figure 3, very few patients exceeded 30 minutes of waiting, which means very few overflowed to the next slot. Secondly, although majority of patients did not have to wait more than 20 minutes, still there are two cases where patients have to wait more than 45 minutes in runs 1 and 2, respectively. There is a high likelihood that these two patients will renege. The reason for this long time is that the objective function considers patient overflow from one slot to the other and does not differentiate between an overflow by a minute or an overflow by a long time period. This can be avoided if the direct waiting time is considered instead of overflow, but this is a non-trivial task due to two main reasons: there is no deterministic relationship that describes waiting time as a function of probabilistic service time and probabilistic show-up behaviour. The situation is more complex if walk-ins and probabilistic arrival are both integrated in the system. Table 5 summarises the results of the dynamic scheduling approach proposed by Muthuraman and Lawely (2008).

In the proposed approach, the patient-dependent approach to treat no-show rates is analysed. Based on historical data available to clinic, patients are classified into three levels on no-show rates (25%, 50%, and 75%). This scenario resulted in the worst overall utility (about 26%) with lowest number treated (14 patients) and lowest utilisation (32%). 16 patients did not show up resulting in zero reneging.

Remedial actions are set to first test the impact of allowing walk-in patients into the clinic. Then, three overbooking (OB) strategies are tested; add (overbook) an extra patient every one hour, every half hour, and every 15 minutes (double booking). The results of remedial actions are summarised in Table 6.

Table 6 Simulation results of patient-dependent classification of no-show rates

	Simulated remedy	*Patients*				*Performance measures*			*Utility*
		WI	*RN*	*NS*	*OB*	*DT*	*WT*	*DU*	*U*
1	Accept walk-in patients	3	0	16	0	17.0	6.2	49.1%	62.82
2	OB (add a patient every one hour)	0	2	16	8	19.0	9.1	57.4%	65.52
3	OB (add a patient every half hour)	0	2	16	16	26.0	18.3	78.9%	73.16
4	OB (every 15 min): *double booking*	0	9	16	32	28.0	38.7	88.6%	63.37

Due to the relatively high no-show rates, allowing walk-ins significantly improved performance by treating three more patients and improving utilisation to 50% with a 63% overall utility. Based on the results, the best overbooking strategy at these no-show rates is to overbook a patient every half-hour (double the average treatment time). This boosts the overall utility to 73% even though 16 patients will not show. Double booking results in lower performance measures and overall utility.

6 Conclusions

DPM provides analysts with a flexible platform to model the operation of healthcare clinics and assess clinics performance measures. It also allows for testing various overbooking strategies to deal with patients' no-show-ups. This paper presented a literature review on the topic and illustrated the value of DPM. Building on the previous research, a simulation approach is proposed to analyse the impacts of patients' no-show rates on clinic performance and to test several overbooking strategies. The key contributions of the proposed approach include using DES for dynamic modelling of clinic performance, assessing the impact of overbooking on clinic revenue (patients treated), capacity (utilisation), customer service (wait time), and using an overall utility function that incorporates the impact of variability in scheduling and treatment times, the impact of reneging, and the impact of a walk-in policy.

Results showed that variability in arriving and treatment times disturbs clinic operation resulting in less patients treated per day with longer waiting time and reduced utilisation. Results also showed that treatment time variability has more negative impact than the variability in patients' inter-arrival times. As the percentage of patients' no-show increases the number of reneged patients decreases and customer service in terms of waiting time improves. However, clinic utilisation and the number of treated patients drop. At a 30% no-show rate, the clinic performance significantly drops to a 55% utilisation and an overall clinic utility of 63%. Patient waiting time and patient reneging inevitably increase with appointment overbooking. Overbooking provides greater utility when no-show rates are high because of the increased improvement opportunity. Patient-based dynamic overbooking with high no-show rates showed that the best overbooking strategy is to overbook a patient periodically at double the treatment time. The clinic manager has to assess the financial situation of the clinic and select the most effective overbooking strategy to improve clinic performance. Reducing variability in service time first and later in patients' arrival times would be a major recommendation in this regard.

The proposed simulation-based overbooking model accommodates a wide range of clinic sizes and no-show rates, allowing its application in a variety of clinical practices. Other solution approaches that can also be evaluated include dynamic overbooking where overbooking rate changes with time of day and the two physicians scheduling alignment. Suggested further work includes applying simulation-based DOE to analyse the overbooking strategies and using simulation-based optimisation search engines such as genetic algorithm (GA) and simulated annealing (SA) to optimise the utility functions of healthcare clinics.

References

Abdus-Salaam, H. and Davis, L. (2010) 'Examining the impact of dependent demand arrivals at a pediatric dental clinic using multivariate statistics', *Proceedings of the 2010 Industrial Engineering Research Conference*, Cancun, Mexico.

Ahmed, M. and Alkhamis. T. (2009) 'Simulation optimization for an emergency department healthcare unit in Kuwait', *European Journal of Operational Research*, Vol. 198, No. 3, pp.936–994.

Banks, J. (Ed.) (1998) *Handbook of Simulation, Principles, Methodology, Advances, Applications, and Practice*, John Wiley & Sons, Inc., New York.

Bech, M. (2005) 'The economics of non-attendance and the expected effect of charging a fine on non-attendees', *Health Policy*, Vol. 74, No. 2, pp.181–191.

Boxerman, S.B. (1996) 'Simulation modeling: a powerful tool for process improvement', *Best Practices and Benchmarking in Healthcare*, Vol. 1, No. 9, pp.109–117.

Brahimi, M. and Worthington, D.J. (1991) 'Queuing models for out-patient appointment systems: a case study', *The Journal of the Operational Research Society*, Vol. 42, No. 3, pp.733–746.

Campbell, J.D., Chez, R.A., Queen, T., Barcelo, A. and Patron, E. (2000) 'The no-show rate in a high-risk obstetric clinic', *Journal of Women Health and Gender-Based Medicine*, Vol. 9, No. 8, pp.891–895.

Cayirli, T.E. and Veral, H.R. (2006) 'Designing appointment systems for ambulatory care services', *Health Care Management Science*, Vol. 9, No. 1, pp.47–58.

Chakraborty, S., Muthuraman, K. and Lawley, M. (2011) 'Sequential clinical scheduling with patient no-shows and general service time distributions', *IIE Transaction*, Vol. 214, No. 3, pp.780–795.

Chung, M.K. (2002) 'Tuning up your patient schedule', *Family Practice Management*, Vol. 9, No. 1, pp.41–48.

Colli, F., Ferreira, R., Almeida, R.M. and Pereira, W. (2007) 'Computer simulation and discrete-event models in the analysis of a mammography clinic patient flow', *Computer Methods and Programs in Biomedicine*, Vol. 87, No. 3, pp.201–207.

Consumer Reports (2007) Consumer Reports News release, February, pp.32–36.

Denton, B. and Gupta, D. (2003) 'A sequential bounding approach for optimal appointment scheduling', *IEE Transaction*, Vol. 35, No. 11, pp.1003–1016.

Deyo, R.A. and Inui, T.S. (1980) 'Dropouts and broken appointments', *Medical Care*, Vol. 18, No. 11, pp.1146–1157.

Dove, H.G. and Schnieder, K.C. (1981) 'The usefulness of patients', *Individual Characteristics in Predicting No-Shows in Outpatient Clinics' Medical Care*, Vol. 19, No. 7, pp.734–740.

Goldman, L., Freidin, R., Cook, E.F., Eigner, J. and Grich, P. (1982) 'A multivariate approach to the prediction of no-sow behavior in a primary care center', *Archives of Internal Medicine*, Vol. 142, No. 3, pp.563–567.

Greasley, A. (2006) 'Using process mapping and business process simulation to support a process-based approach to change in a public sector organization', *Technovation*, Vol. 26, No. 1, pp.95–103.

Gupta, D. and Denton, B. (2008) 'Appointment scheduling in health care: challenges and opportunities', *IIE Transactions*, Vol. 40, No. 9, pp.800–819.

Hamilton, W., Round, A. and Sharp, D. (1999) 'Effect on hospital attendance rates of giving patients: a copy of their referral letter: randomized controlled trial', *Bio-Medical Journal*, Vol. 318, No. 7195, pp.1392–1395.

Harper, P. and Gamlin, H. (2003) 'Reduced outpatient waiting times with improved appointment scheduling: a simulation approach', *OR Spectrum*, Vol. 25, No. 2, pp.207–222.

iGrafx process 2011 for Six Sigma (www.igrafx.com).

Jacobson, S.H., Hall, S.N. and Swisher, J.R. (2006) 'Discrete-event simulation of health care systems. Chapter 8 in patient flow: reducing delay in healthcare delivery', Hall, R.W. (Ed.): *International Series in Operations Research & Management Science*, Vol. 91, pp.211–252, Springer.

Jain, Sh. and Chou, C. (2000) 'Use of an orientation clinic to reduce failed new patient appointments in primary care', *Journal of General Medicine*, Vol. 15, No. 12, pp.878–880.

Jun, J., Jacobson, S. and Swisher, J. (1999) 'Application of discrete-event simulation in health care clinics: a survey', *Journal of the Operational Research Society*, Vol. 50, No. 2, pp.109–123.

Kennedy, J. and Hsu, J. (2003) 'Implementation of an open access scheduling system in a residency training program', *Family Medicine*, Vol. 35, No. 9, pp.666–670.

Kheirkhah, P., Feng, Q., Sharafkhaneh, A. and Obermeier, L. (2009) 'A statistical and economical analysis of no-show rates in a VA hospital', *Proceedings of the 2009 IE Conference*, Miami, FL.

Kim, S. and Giachetti, R.E. (2006) ' A stochastic mathematical appointment overbooking model for healthcare providers to improve profits', *IEEE Transactions on Systems, Man, and Cybernetics*, Vol. 36, No. 6, pp.1211–1219.

Kosfeld, M. (1998) 'Warehouse design through dynamic simulation', *Proceedings of the 1998 Winter Simulation Conference*.

Lacy, N.L., Paulman, A., Reuter, M.D. and Lovejoy, B. (2004) 'Why we don't come: patient perceptions of no-show', *Annals of Family Medicine*, Vol. 2, No. 6, pp.541–545.

LaGanga, L. and Lawrence, S. (2007) 'Clinic overbooking to improve patient access and increase provider productivity', *Decision Sciences*, Vol. 38, No. 2, pp.251–276.

Law, A. (2006) *Simulation Modeling and Analysis*, 4th ed., McGraw-Hill College, Columbus, OH.

Moore, C.G. and Wilson-Witherspoon, P. (2001) 'Time and money: effects of no-shows at a family practice residency clinic', *Family Medicine*, Vol. 33, No. 7, pp.522–527.

Muthurman, K. and Lawley, M. (2008) 'A stochastic overbooking model for outpatient clinical scheduling with no-show', *IIE Transaction*, Vol. 40, No. 9, pp.820–837.

Nelson, B.L., Banks, J., Carson, J.S. and Nicol, D.M. (2006) *Discrete-Event System Simulation*, 3rd ed., Prentice Hall, Upper-Saddle, NJ.

Penneys, N. and Glaser, D.A. (1999) 'the incidence of cancellation and non-attendance at a dermatology clinic', *Journal of American Academy of Dermatology*, Vol. 40, No. 5, pp.714–718.

Robinson, L.W. and Chen, R.R. (2003) 'Scheduling doctors' appointments: optimal and empirically-based heuristic policies', *IIE Transactions on Scheduling and Logistics*, Vol. 35, No. 3, pp.295–307.

Rohleder, T.R. and Klassen, K.J. (2006) 'Rolling horizon appointment scheduling: a simulation study', *Health Care Management Science*, Vol. 5, No. 3, pp.201–209.

Satianti, Bh., Miller, S. and Patel, D. (2009) 'No-show rates in the vascular laboratory: analysis and possible solutions', *Journal of Vascular Intern Radiology*, Vol. 20, No. 1, pp.87–91.

Vemeulen, I., Bohte, S., Elkhuizen, S., Lameris, H., Bakker, P. and Poutre, H. (2009) 'Adaptive resource allocation for efficient patient scheduling', *Artificial Intelligence in Medicine*, Vol. 46, No. 1, pp.67–80.

Woodcock, E. (2003) *Mastering Patient Flow: More Ideas to Increase Efficiency and Earnings*, MGMA, Englewood, CO.

Zeng, B., Zhao, H. and Lawley, M. (2009) 'Clinic overbooking and patient responses: a game theoretical approach', *Proceedings of the 2009 Industrial Engineering Research Conference*, Miami, Fl.