DECISION SCIENCES
A JOURNAL OF THE DECISION SCIENCES INSTITUTE

*Decision Sciences*
*Volume 45 Number 5*
*October 2014*



# Strategies for Appointment Policy Design with Patient Unpunctuality*

Kenneth J. Klassen and Reena Yoogalingam†
*Brock University, 500 Glenridge Ave, St Catharines, ON L2S 3A1, Canada,*
*e-mail: kklassen@brocku.ca, ryoogalingam@brocku.ca*

## ABSTRACT

Appointment policy design is complicated by patients who arrive earlier or later than their scheduled appointment time. This article considers the design of scheduling rules in the presence of patient unpunctuality and how they are impacted by various environmental factors. A simulation optimization framework is used to determine how to improve performance by adjusting the schedule of appointments. Prior studies (that did not include patient unpunctuality) have found that a scheduling policy with relatively consistent appointment interval lengths in the form of a dome or plateau dome rule to perform well in a variety of clinic environments. These rules still perform reasonably well here, but it is shown that a combination of variable-length intervals and block scheduling are better at mitigating the effects of patient unpunctuality. In addition, performance improves if the use of this policy increases toward the end of the scheduling session. Survey and observational data collected at multiple outpatient clinics are used to add realism to the input parameters and develop practical guidelines for appointment policy decision making. [Submitted: October 12, 2012. Revised: June 26, 2013. Accepted: July 12, 2013.]

***Subject Areas: Appointment Scheduling, Simulation Optimization, and Patient Unpunctuality.***

## INTRODUCTION

Patient unpunctuality is prevalent in many healthcare settings and is particularly noticeable in outpatient clinics. The design of an efficient appointment policy is complicated by patients who arrive earlier or later than their scheduled appointments. Late patients may result in physician idleness, which represents lower utilization and productivity. They may also extend waiting for subsequent patients and increase overtime for the clinic. However, refusing to serve the patient (i.e., rescheduling them for another day) may not be desirable because this would result in reduced utilization of the clinic. Early patients also pose a problem because they contribute to congestion in the waiting room which can lead to patient dissatisfaction and issues associated with staff morale (Rohleder, Lewkonia, Bischak, Duffy, & Hendijani, 2011). Furthermore, early patient arrivals are difficult to control

*This research was supported in part by Canada's Natural Sciences and Engineering Research Council (NSERC), Discovery Grant #283333.

†Corresponding author.

because it is not usually socially acceptable to "penalize" a patient for being early. Consequently, it is worthwhile to design appointment policies which mitigate the effects of both early and late arrivals.

Patient unpunctuality is defined as the difference between a patient's appointment time and their arrival time, indicating they can arrive either early or late for their scheduled appointment. Most empirical evidence indicates that, on average, patients arrive early for appointments (Fetter & Thompson, 1966; Cox, Birchall, & Wong, 1985; Brahimi & Worthington, 1991; Klassen & Rohleder, 1996). However, there is some degree of variability associated with the arrival process. This variability is cause for concern because it can be a major cause of poor system performance including long facility idle times and waiting room congestion (Blanco White & Pike, 1964; Cayirli & Veral, 2003; Tai and Williams 2011). Therefore, it is important to incorporate this factor when modeling appointment systems to achieve improvements in operational efficiency from a scheduling perspective.

The problem of appointment scheduling is typically one of minimizing the weighted cost of patient waiting and physician or clinic idle time and/or overtime. Many prior studies have assumed that patients arrive on time for their scheduled appointments (e.g., Denton & Gupta, 2003; Robinson & Chen, 2003). In such cases, patients are seen by the available physician in order of their appointment. When patients are assumed to arrive unpunctually, the order in which patients are seen becomes an issue. For example, in practice, it is common to move up an early arrival when a patient is late (Powers, 2011). This aspect has not received much attention in the literature. A notable exception is Cayirli, Veral, and Rosen (2006, 2008) where late patients are penalized in direct proportion to their lateness. In this case, the late patient would lose one place in the queue for each amount of time equal to the mean service time they are late.

This article addresses the design of effective scheduling policies in the presence of patient unpunctuality for a single-server, single-stage outpatient scheduling environment. We consider the best variable interval appointment scheduling policies and the robustness of the findings are tested by including factors such as patient no-shows, physician lateness, physician interruptions, and different valuations of the physician's time in relation to the patients' time. The problem environment and parameter values are based on survey and observational data collected at multiple outpatient clinics and prior empirical studies. A simulation optimization approach is used which searches for the good solutions while accounting for the complex environmental characteristics and stochastic factors present in the model. As far as can be determined, these factors have not been simultaneously studied in the literature. A goal is to develop a set of general rules and principles that can be used to mitigate the effects of unpunctuality in a variety of clinic environments.

The article is organized as follows. In the next section a review of the related literature is provided. Then, an overview of the data collected for this study and the problem formulation are discussed. Finally, results and analysis are presented, ending with a discussion and managerial implications.

## RELATED LITERATURE

Patient unpunctuality has been examined in several simulation studies. These show that patient unpunctuality has a negative impact on the performance of the

system and propose scheduling rules to mitigate its effects. Scheduling rules are generally defined by two components: the number of patients assigned to each appointment interval or "block" and the length of the appointment interval. For example, a single-block fixed interval scheduling rule has one patient assigned to each appointment interval and all intervals of equal length. A multiblock variable interval rule has two or more patients assigned to each appointment interval which vary in length. Most simulation studies consider fixed interval policies of single block and/or multiple blocks (e.g., Bailey, 1952; LaGanga & Lawrence, 2007; Cayirli et al., 2008). Bailey (1952) designed a rule that has been shown effective in many environments; a fixed interval rule where two patients are placed in the first interval and then the rest in individual blocks.

Analytical studies typically result in single-block, variable interval rules (e.g., Denton & Gupta, 2003). Recent studies have found a "dome" rule to maximize performance of appointment systems under various environmental conditions (e.g., Wang, 1997; Denton & Gupta, 2003; Robinson & Chen, 2003; Kaandorp & Koole, 2007). The dome rule is a variable interval, single block rule where appointment intervals gradually increase in length toward the middle of the session, then gradually decrease in length toward the end of the session. More recently, a simulation-optimization approach has been used to study the problem. These studies have found variable interval rules with both single and multiple blocks to perform well for a wide variety of clinic environments (e.g., Klassen & Yoogalingam, 2008, 2009). For a comprehensive review and commentary on studies where patient arrive punctually, the reader is referred to Cayirli and Veral (2003) and Gupta and Denton (2008). In this study, a simulation optimization model based on empirical data is used to determine the best scheduling rules to mitigate the effects of patient unpunctuality.

Blanco White and Pike (1964) found that patients are on average punctual, although the variance in arrival time is high. After modeling the two specific empirical patterns they observed (where average unpunctuality is zero), they found that only multiblock (fixed interval, blocks of three) rules performed well in alleviating the effects of patient unpunctuality. Based on this, they suggest that if patients are actually early on average, the initial block of patients should be reduced in size. Cox, Birchall, and Wong (1985) studied two clinics and also found that multiblock systems (blocks of three) work well, although in one clinic a single-block system performed just as well. Fetter and Thompson (1966) found empirically that approximately 81% of patients arrive early and that the average earliness for these outpatients was 18.4 minutes in one hospital and 17.2 minutes in the other. They model unpunctuality, but do not propose scheduling rules; instead they report that if patients are taught to arrive on time ($\pm$5 minutes), waiting time would be substantially reduced. Vissers (1979) studied cases with a low level of earliness (5 minutes) and finds that either a single block system with an initial block of two patients (i.e., Bailey' rule), blocks of three patients, or the physician arriving late can have similar effects in improving performance. The studies mentioned above all measured waiting time from the time of patient arrival.

More recently, Cayirli, Veral, and Rosen (2006, 2008) studied a variety of factors to determine their impact on scheduling policy. Their studies found patient unpunctuality to be one of the more important factors. They showed that as patient earliness increases, physician idle time, overtime, and even patient waiting time

improve. The authors did not include the cost of waiting before the scheduled appointment time and focus on patient classification and sequencing rules and thus, their results vary from the papers discussed earlier. They propose several rules for consideration by a decision maker depending on the environmental conditions faced by the clinic. For example, Bailey's rule, a multiblock fixed interval rule (blocks of two patients every second appointment interval), and an individual block fixed interval rule perform well in minimizing costs to the clinic. Although patient unpunctuality impacted the results, this study did not isolate the impact that patient unpunctuality has on appointment policy design.

Some studies have also focused on how to model patient unpunctuality. Most recent studies use empirical data to determine an appropriate distribution to represent this factor. In their two separate samples of 91 and 134 patients, Alexopoulos, Goldsman, Fontanesi, Kopald, and Wilson (2008) found that the family of Johnson distributions provided a good fit for unpunctuality but do not attempt to determine a schedule to minimize its impact. One of their datasets had patients arriving early on average (a mean of –3.49 minutes and standard deviation of 19.51), whereas the other had late arrivals on average (a mean of 8.88 minutes and standard deviation of 18.67). As far as can be determined, all other literature reports that arrivals are either on-time or early (on average). Other theoretical distributions for patient unpunctuality that have been considered include the normal distribution (Cayirli et al., 2006, 2008), exponential for late patients (Cox et al., 1985), and Pearson type VII (Blanco White & Pike, 1964). Most data suggests that unpunctuality is independent of patient appointment time (e.g., Blanco White & Pike, 1964; Cayirli et al., 2006). More recently, the work by Robinson and Chen (2011) provides a pragmatic method for calculating the actual cost of patient waiting.

Patient waiting time has been measured from the time of the patient's appointment (e.g., Cayirli et al., 2006, 2008) and from the time of the patient's arrival (e.g., Blanco White & Pike, 1964; Fetter & Thompson, 1966; Vissers, 1979; Cox et al., 1985). Insights can be gained from both approaches and both are examined in this article. As reviewed above, empirical data suggests that for most clinics patients arrive early on average. From the clinic's perspective, it is easy to argue that a patient's arrival time is within their control and any waiting prior to appointment time is not a relevant cost to the clinic (Maister, 1985). However, one of the main objectives of appointment systems is to provide patients with a higher quality of service (Liu, Ziya, & Kulkarni, 2010). Janakiraman, Meyer, and Hoch (2011) find that present and future business is reduced if there is congestion in the waiting area (through balking, reneging, and less chance of returning in the future). If early arrivals are not accounted for, this could have a negative impact on the perception of service quality. For example, a crowded waiting area could be perceived as the physician being behind schedule and an implied longer waiting time.

## DATA AND PROBLEM ENVIRONMENT

### Data

Data was collected by observation at two outpatient clinics, resulting in 570 patients observed across 35 sessions. In addition, physicians and a senior staff member

**Table 1:** Comparison of nurse and physician responses and to data from observation.

| Question | Nurse Average[a] | Physician Average[a] | Clinic 1 Observed[b] | Clinic 2 Observed[b] |
|---|---|---|---|---|
| Number of patients served in a day (on average) | 28.21 | 30.73 | 28.92 | 33.9 |
| Percentage of patient assessments that start: | | | | |
| *Before the time of their appointment (> 2 minutes before)* | 4.9% | 5.4% | 46.5% | 25.4% |
| *At the time of their appointment (± 2 minutes)* | 33.3% | 37.3% | 13.7% | 13.5% |
| *After the time of their appointment (> 2 minutes after)* | 61.8% | 57.4% | 39.8% | 61.2% |
| Reason for late start times: | | | | |
| *Patients are late* | 10.8% | 9.7% | 9.22% | 0.53% |
| *Physician falls behind schedule* | 89.2% | 90.3% | 90.78% | 99.47% |
| Average physician lateness in starting sessions | 0.00 | 0.00 | 0.83 | 7.21 |

[a]Surveys done at 10 different offices. Note: although it is desirable to compare the responses of the nurse & physician at the observed clinics directly to their own observed data, ethical requirements did not allow us to isolate individual surveys.
[b]Sample size for all is 570, except physician lateness which is 35.

involved with day-by-day operations (usually the nurse) at 10 different medical offices filled out surveys that represent their impressions of clinic operations. Because it was possible to observe exactly when the physician entered and left the examination room (i.e., not just when the patient entered), patient waiting includes time in the waiting room and waiting in the examination room. The goal was not to perform a case study of the clinics but rather to gather information to motivate the problem environment and input parameters of the study.

A comparison of some key aspects of how the survey data corresponds to the observed data is provided in Table 1. A high level of detail was collected in regard to patient unpunctuality and late appointment start times, including reasons why appointments start late.

The observed data in Table 1 shows that a large number of appointments start before their scheduled start time and the survey data suggests that physicians and their nurses are generally not aware of this. It appears that physicians and nurses consider these early starts to be "on-time." There is fairly strong correspondence between the observed and survey data in terms of the number of appointments that start late and the reasons. In most cases, it is because the doctor is falling behind. The data also show that nurses and physicians perceive that sessions start on time and that is true especially for Clinic 1. Clinic 2 starts approximately 7 minutes late on average. Many other statistics were collected including customer no-shows and physician interruptions. Clinics 1 and 2 experienced no-shows at 6.63% and 7.72%, respectively. Doctors may experience interruptions to patient care due to things such as phone calls, charts to write up, and consultation with staff or other physicians (Rising, Baron, & Averill, 1973). The data showed that 7.86%

of appointments in Clinic 1 were interrupted compared to 11.95% of appointments in Clinic 2. In the vast majority of prior studies, interruptions have been ignored, suggesting their duration is incorporated into the appointment times. It has recently been shown that this is a valid approach. Klassen and Yoogalingam (2008) showed that if interruptions are modeled such that they are included in the appointment interval, then the scheduling rule is not affected. However, if they are not accounted for during an appointment, then the scheduling rule has to be adjusted by increasing the interval lengths (Klassen & Yoogalingam, 2013).

The observed data was fit to distributions based on the chi-squared test ($\alpha = 0.05$). Patient unpunctuality (the difference between a patient's arrival and appointment time) followed a normal distribution with a mean of –9.39 minutes and standard deviation of 13.7. Physician lateness (in starting the session) was also normally distributed with a mean of 2.62 minutes and standard deviation of 5.49. The service duration followed a lognormal distribution with a mean of 5.41 minutes and standard deviation of 3.52. Compared to data collected for prior studies, these physicians are more punctual and the service times are shorter. The data for patient unpunctuality, suggests that although patients arrive early on average, there is considerable variation in their arrival times. In addition, we examined the data to see if arrival patterns varied throughout the session but found unpunctuality levels to be consistent across the session. This is consistent with prior studies which have found that unpunctuality is independent of appointment time (e.g., Blanco White & Pike, 1964; Cox et al., 1985; Cayirli et al., 2006).

## Problem Formulation

The model consists of a single server in an outpatient clinic over the course of a session. The problem can be seen as one of determining the start time of each of $N$ appointments, $t_i$, where $i = 1, 2, \ldots, N$. For simplicity and bevause the identity of the patient is not significant in the calculation of waiting time, we present the objective function as the sum of patient waiting time and doctor idle time based on the order in which patients are served. Accordingly, $PW_j$ is the waiting time of the $j$th person seen by the system and $DI_j$ is the idle time associated with the $j$th person served.

$$\min \text{PWDI} = c_{pw} E\left(\sum_{j=1}^{N} PW_j\right) + c_{di} E\left(\sum_{j=1}^{N} DI_j\right), \tag{1}$$

s.t.

$$t_1 \geq 0 \tag{2}$$

$$t_N \leq T \tag{3}$$

$$t_1 \leq t_2 \leq \ldots \leq t_N \tag{4}$$

$$t_i \text{integer} \tag{5}$$

PWDI is the expected total cost of patient waiting time (PW) and doctor idle time (DI). The cost coefficients for patient waiting time and doctor idle time are denoted by $c_{pw}$ and $c_{di}$, respectively. It is assumed that no appointment is scheduled beyond the planned session end time, $T$. The appointment start times are constrained to take on integer values for this study in order to more closely mimic reality (Klassen & Yoogalingam, 2009; Begen & Queyranne, 2011). In this study, the waiting time of patients depends on their arrival time as well as the arrival and service durations of prior patients, and doctor lateness ($\lambda$). Given that the order in which patients are served may differ from the order in which they are assigned their appointments, the following notation is used. The arrival time of the $j$th person served is denoted by $A_j$ and their service duration by $s_j$. The start of service for the $j$th person served is denoted by $SS_j$ and their service end time by $SE_j$, where $SE_j = SS_j + s_j$. Then the waiting time of the $j$th person served is given by

$$PW_1 = \max\{\lambda - A_1,\ 0\}$$

$$PW_j = \max\{SE_{j-1}, A_j\} - A_j \; j = 2, ..., N \tag{6}$$

Each patient is seen in the order of their appointment unless a patient is late and/or the doctor is idle and there is another patient waiting (i.e., one that arrived early). For example, if the patient assigned to appointment $i$ arrives before patient assigned to appointment $i-1$ and there is no one in the queue and the physician is available, the patient assigned to appointment $i$ will be seen before the patient assigned to appointment $i-1$. Following some prior studies (e.g., Blanco White & Pike 1964; Fetter & Thompson, 1966; Vissers, 1979; Cox et al., 1985), patient waiting time is measured as the difference between the patient's arrival time and service start time to capture the variability in the system and costs of patient waiting resulting from early arrivals.

Doctor idle time is a function of both the doctor's lateness and patients' unpunctuality.

$$DI_1 = \begin{cases} \max\{A_1 - t_1, 0\} \text{ if} \lambda = 0 \\ \max\{A_1 - \lambda, 0\} \text{ if} \lambda > 0 \end{cases} \tag{7}$$

$$DI_j = \max\{A_j - (A_{j-1} + PW_{j-1} + s_{j-1}), 0\},\ j = 2, ..., N \tag{8}$$

The session end time is denoted by $SET$ and doctor overtime by $DO$ to determine the impact of patient unpunctuality on overtime. It is assumed all appointments will be seen regardless of the degree of unpunctuality.

$$SET = \max\{A_N, A_{N-1} + PW_{N-1} + s_{N-1}\} + s_N \tag{9}$$

$$DO = \max\{0, SET - T\} \tag{10}$$

Many studies have included overtime as a performance measure (Vanden Bosch et al., 1999; Denton & Gupta, 2003). It is possible that if overtime instead of idle time is used in the performance measure (PWDO) or added to the performance measure (PWDIDO), the resulting scheduling policies would result in less overtime. Doctor overtime usually does not represent wages, but rather the cost to keep the clinic open longer and the resulting negative impact on doctors and staff. In

order to test whether including DO would affect the best scheduling policy and/or result in reduced overtime levels when compared to using (1), the following secondary measures were used, where $c_{do}$ is the cost coefficient for doctor overtime:

$$\min \text{PWDO} = c_{pw} E\left(\sum_{j=1}^{N} PW_j\right) + c_{do} E\,(DO) \tag{11}$$

$$\min \text{PWDIDO} = c_{pw} E\left(\sum_{j=1}^{N} PW_j\right) + c_{di} E\left(\sum_{j=1}^{N} DI_j\right) + c_{do} E\,(DO) \tag{12}$$

## Experimental Design and Other Factors

In this study, a 4-hour session is modeled. Similar to prior research (e.g., Klassen & Yoogalingam, 2009) the distribution for patient service durations is modeled as lognormal with a mean of 10 minutes and a standard deviation following a lognormal distribution with a mean and standard deviation of 6. Modeling the distribution of service durations in this way allows the model to capture the relatively common 10-minute appointment length found in practice as well as differences in patient variability distributions. This results in a clinic scheduling 24 appointments per session.

Patient unpunctuality is modeled at zero and using a normal distribution with a mean of 0, –10, and –20 minutes and standard deviations of 15 and 30. This distribution is consistent with the data collected for this and other studies (i.e., patients are early on average). Patient waiting time is also measured using two different metrics: PW, where total waiting time is used in the objective function and PW($t_i$), where only waiting from scheduled appointment time is used in the objective function.

Several factors were tested along with patient unpunctuality in order to determine if their presence had an impact on the results and whether a different scheduling policy performs better when they are included. These are physician lateness, patient no-shows, physician interruptions, the cost of the clinic's time in relation to the patients' time, and how waiting time is measured.

Physician lateness has been shown to be an important factor. Liu and Liu (1998) observe that physicians often arrive late, although the lateness never exceeded 6 minutes in that study. Due to this minimal lateness, their experiment with a model of a multiserver clinic shows that lateness seems to have little effect on the results. Blanco White and Pike (1964) consider lateness up to 10 minutes in their simulation study and find that patient waiting times are sensitive to physician lateness. They propose reducing the number of patients seen in a particular session to compensate for the lateness of the physician. Fetter and Thompson (1966) also find through simulation that the degree of lateness of the physician is a major determinant of patient waiting time, although it reduces idle time. Due to lack of data, modeling lateness has often been restricted to distributions such as the uniform or a constant (e.g., Blanco White & Pike, 1964; Fetter & Thompson, 1966; Babes & Sharma, 1991; Liu & Liu, 1998; Klassen & Yoogalingam, 2013). In this study this

factor was modeled using a normal distribution and tested with a mean at three levels of 0, 10, and 20 minutes and a standard deviation of 5 which is consistent with our observed data and prior studies.

The impact of patient no-shows is also studied. No-shows commonly occur in outpatient clinic environments and can be considered to be an extreme form of patient unpunctuality (lateness) (Kaandorp & Koole, 2007). It is assumed that each patient has the same probability of not showing up for their appointment. Probabilities are modeled at three levels: 0%, 12.5%, and 25%. These reflect the range found in a variety of practices (Cayirli & Veral, 2003). Given the percentage of no-shows, the appropriate number of patients are then overbooked so that utilization remains the same on average, enabling performance of different scenarios to be compared directly.

Physician interruptions have received some attention in the literature as a factor that impacts appointment policy design. Patient waiting times may increase when there are phone calls to deal with, charts to write up, and consultation with staff or other physicians (Rising et al., 1973). Interruptions are typically defined as occurring between patient appointments and Klassen and Yoogalingam (2008) found empirically that interruptions between appointments are common (occurring for 22.65% appointments). This factor is modeled at two levels: 0% and 20%.

To capture the relative cost to the clinic of physician idle time to patient waiting time, the coefficients for the cost of doctor idle time ($c_{di}$) in Equation (1) and overtime in Equation (9) were tested. Prior studies have tested a range of values from one to nine (Denton & Gupta, 2003) and values as high as 100 (Robinson & Chen, 2003; Cayirli et al., 2008). In this study, values of 1, 5, 9, 20 and 50 (with $c_{pw}$ at 1) are tested in order to study the changes to the system as the physician's time is weighted more heavily.

## The Simulation Optimization Algorithm

The multiple sources of randomness and distributions present in this study require the use of a technique that is capable of accounting for the stochastic parameters of the problem and finding good solutions in a large, complex search domain. Simulation optimization is a stochastic optimization technique that iteratively generates sets of decision variable values that are evaluated for performance by a simulation module. Because the objective function and constraints in this study are stochastic functions, the best solution for the problem is a statistic (e.g., mean) of the resulting objective function values from each iteration. This method is well suited to problems where closed-form solutions are not available for the problem and have to be estimated (in this case by simulation) (Andradóttir, 2006).The simulation optimization problem can generally be defined (Fu, 2002; Andradóttir, 2006) as minimizing the following objective function

$$\min f(\theta) = E[\gamma(\theta, \omega)], \tag{13}$$

$$\text{s.t.} \theta \in \Theta$$

where $\gamma$ is the sample performance measure, $\theta$ is the vector of input variables, and $\omega$ represents a simulation replication. Accordingly, the problem is one of

determining the vector of appointment times or scheduling policy, $\theta$, that minimizes the performance measure PWDI where the vector $\omega$ corresponds to one realization of the vector or arrival times, $A_i$, service times, $s_j$, and physician lateness, $\gamma$. In this study, the simulation optimization procedure embedded in OptQuest (Laguna, 1997) is used. This procedure combines scatter search and tabu search heuristics, which have been shown to be capable of finding good solutions for appointment scheduling problems (Klassen & Yoogalingam, 2009).

Scatter search is a population-based search heuristic which exploits information derived from prior solutions in its search. The algorithm begins by generating an initial population of candidate solutions. Each candidate solution is an $n$-vector of input factors or decision variable values, $\boldsymbol{\theta}$. An upper and lower bound for each input factor, $u_i$ and $l_i$, respectively, is specified. Given this, one initial solution vector will consist of components

$$\theta_i = l_i + \frac{(u_i - l_i)}{2} \text{for } i = 1, 2, \ldots, n \tag{14}$$

Additional points are generated for the initial population using a diversification generation method with a seed solution vector (e.g., [0, 0, ..., 0]). The aim is to generate an initial population that is diverse in the sense that solutions are significantly different from one another. This set of solutions is then evaluated by a simulation module.

This approach creates new solutions by generating linear combinations of prior solutions and a mix of high quality and ordinary solutions to maintain diversity of the population set. It has the distinguishing feature of mapping any infeasible solutions to a feasible one (using a mixed-integer linear programming technique to minimize the absolute deviation between the two points) to optimize the search process. In addition, tabu search is used as a restarting mechanism to ensure a diverse solution set is maintained. Because this approach can be computationally demanding, a neural network accelerator determines the number of replications needed to satisfy the required tolerance. This prevents the algorithm from evaluating solutions it predicts will be inferior to the best solution found.

## RESULTS AND ANALYSIS

Experiments were designed to determine the impact the various factors have on performance and scheduling policy. Each solution was replicated 10,000 times in order to generate statistics for the performance measures. This resulted in 95% confidence intervals across all performance measures (*PW*, *DI*, *DO*, PWDI, PWDO) of 1.26% on average, with a maximum of 3.10%.

Initially, PWDI is used as the performance measure for an experiment with patient unpunctuality (PU) at zero and following a normal distribution with a mean of 0, –10, and –20 minutes, standard deviations of 15 and 30, and $c_{di}$ values of 1, 5, and 9. Figure 1 shows the best solutions found by the simulation optimization algorithm for each level of patient unpunctuality, $c_{di}$, and mean values for the sum of patient waiting time, doctor idle time, and doctor overtime. These results are also summarized in Table 2 where doctor lateness equals zero. Figure 2 illustrates the frequency distribution for PWDI for mean patient unpunctuality levels of –10

**Figure 1:** Effects of patient unpunctuality.

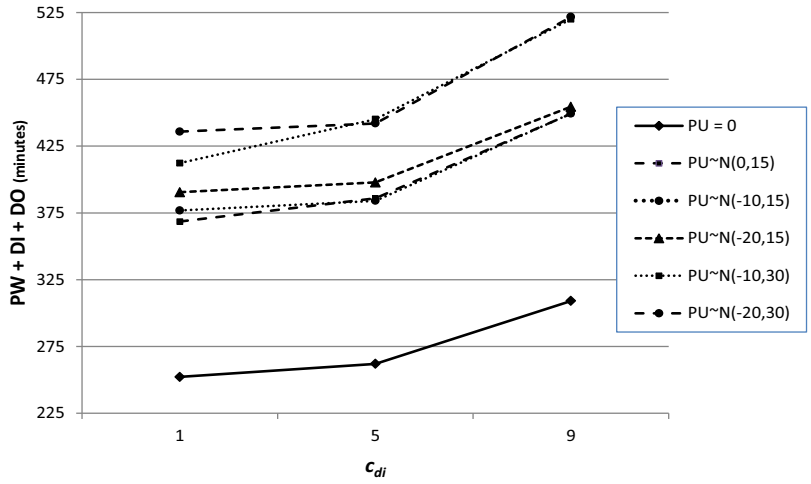
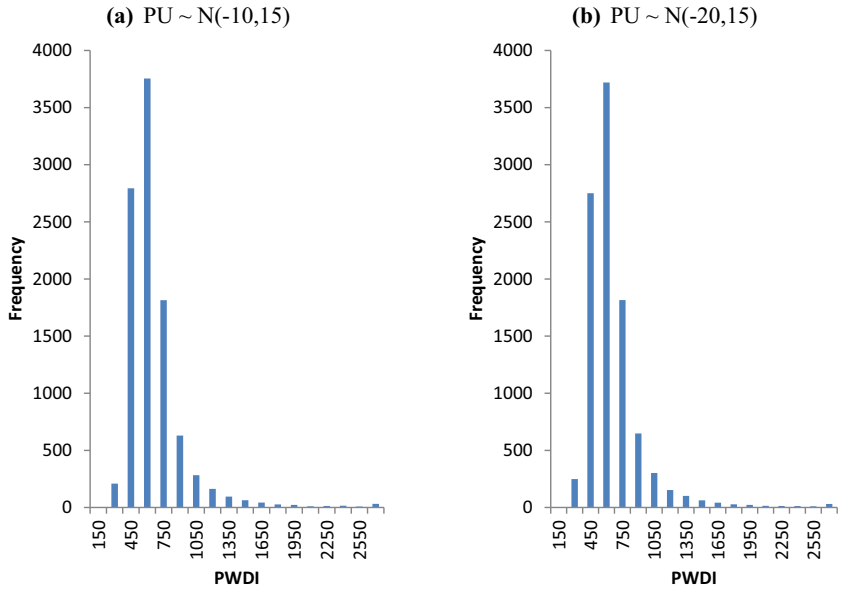



**Table 2:** Comparison of mean performance measures for patient unpunctuality.

| | | $DU = 0$ | | | DU~$N$(20,5) | | |
|---|---|---|---|---|---|---|---|
| | | $c_{di} = 1$ | $c_{di} = 5$ | $c_{di} = 9$ | $c_{di} = 1$ | $c_{di} = 5$ | $c_{di} = 9$ |
| | *PW* | 180.47 | 209.40 | 274.49 | 245.99 | 267.39 | 306.88 |
| $PU = 0$ | *DI* | 36.30 | 26.73 | 17.22 | 28.67 | 21.67 | 16.51 |
| | *DO* | 35.43 | 25.81 | 17.36 | 47.83 | 40.78 | 35.49 |
| | PWDI | 216.77 | 343.07 | 429.45 | 274.66 | 375.73 | 455.46 |
| | *PW* | 260.91 | 316.98 | 407.79 | 329.15 | 353.65 | 430.29 |
| $PU$~$N$(0,15) | *DI* | 54.19 | 34.87 | 20.80 | 43.23 | 32.01 | 21.96 |
| | *DO* | 53.41 | 33.99 | 20.71 | 62.32 | 51.12 | 40.92 |
| | PWDI | 315.10 | 491.34 | 594.95 | 372.37 | 513.68 | 627.90 |
| | *PW* | 286.77 | 315.82 | 408.37 | 372.78 | 400.53 | 445.24 |
| $PU$~$N$(–10,15) | *DI* | 45.43 | 34.53 | 20.52 | 33.42 | 25.75 | 20.52 |
| | *DO* | 44.62 | 33.66 | 20.52 | 52.56 | 44.86 | 39.49 |
| | PWDI | 332.20 | 488.49 | 593.04 | 406.20 | 529.30 | 629.94 |
| | *PW* | 316.52 | 338.29 | 414.47 | 399.86 | 430.44 | 468.03 |
| $PU$~$N$(–20,15) | *DI* | 37.41 | 30.18 | 19.91 | 32.68 | 21.88 | 17.99 |
| | *DO* | 36.55 | 29.35 | 20.05 | 51.82 | 40.99 | 36.97 |
| | PWDI | 353.93 | 489.19 | 593.63 | 432.53 | 539.82 | 629.94 |
| | *PW* | 288.55 | 366.29 | 473.97 | 363.88 | 391.75 | 508.98 |
| $PU$~$N$(–10,30) | *DI* | 63.13 | 39.81 | 22.88 | 47.65 | 39.84 | 23.88 |
| | *DO* | 60.65 | 39.04 | 23.22 | 66.32 | 58.52 | 42.60 |
| | PWDI | 351.68 | 565.32 | 679.87 | 411.53 | 590.98 | 723.90 |
| | *PW* | 314.41 | 360.96 | 475.89 | 408.61 | 411.61 | 506.64 |
| $PU$~$N$(–20,30) | *DI* | 61.97 | 40.87 | 22.71 | 38.81 | 36.11 | 22.97 |
| | *DO* | 59.50 | 40.13 | 23.11 | 57.25 | 54.55 | 41.44 |
| | PWDI | 376.39 | 565.32 | 680.31 | 447.42 | 592.15 | 713.40 |

**Figure 2:** Frequency distribution for PWDI ($c_{it} = 9$).




and –20 minutes and standard deviation of 15 ($c_{it} = 9$) for the best vector of appointment times ($\theta$) found.

The results show that $PW + DI + DO$ increases when patient unpunctuality is present and is higher for higher valuations of doctor idle time. When the standard deviation of patient unpunctuality increases from 0 to 15, this value is 46.21% worse when averaged across all levels of $c_{di}$. When combined with a change in mean unpunctuality to –10 and –20 minutes, $PW + DI + DO$, on average, deteriorates by 47.01% and 50.96%, respectively.

The results in Figure 1 and Table 3 provide two main insights in regard to patient unpunctuality. First, when compared to zero unpunctuality, adding uncertainty (i.e., changing patient unpunctuality from 0 to $N(0,15)$) results in a considerable deterioration of performance. Averaging across all levels of $c_{di}$, performance worsens (i.e., increases) by 48.36% for PW, 36.90% for DI, and 37.54% for DO. PWDI also deteriorates by 41.66%. This is also evident for mean patient unpunctuality levels of –10 and –20 minutes where an increase in the standard deviation from 15 to 30 results in PWDI deteriorating by 12.95% and 12.86%, respectively. Notably, when mean unpunctuality is set at –20, the IT and OT measures become considerably worse when the standard deviation increases from 15 to 30. This demonstrates that uncertain patient arrivals are difficult to adjust for even if patients are on average punctual.

The second insight relates to a change in mean unpunctuality while the standard deviation remains at 15 (i.e., mean changes from 0 to –10 minutes and from 0 to –20 minutes). The sum of patient waiting time, doctor idle time, and overtime

**Table 3:** Percentage change in PWDI for different distributions for patient unpunctuality.

| | Change in PWDI[a] | | | | |
|---|---|---|---|---|---|
| Distribution for *PU* | *N*(0,15) | *N*(–10,15) | *N*(–20,15) | *N*(–10,30) | *N*(–20,30) |
| 0 | 41.66% | 42.90% | 45.23% | 61.41% | 63.91% |
| *N*(0,15) | – | 0.88% | 2.52% | 12.24% | 15.71% |
| *N*(–10,15) | – | – | 1.63% | 12.95% | 14.70% |
| *N*(–20,15) | – | – | – | 11.14% | 12.86% |
| *N*(–10,30) | – | – | – | – | 1.55% |
| | Change in *PW* | | | | |
| 0 | 48.36% | 52.17% | 60.95% | 69.91% | 73.29% |
| *N*(0,15) | – | 2.56% | 8.48% | 12.68% | 16.80% |
| *N*(–10,15) | – | – | 5.77% | 11.66% | 13.88% |
| *N*(–20,15) | – | – | – | 5.57% | 7.67% |
| *N*(–10,30) | – | – | – | – | 1.99% |
| | Change in *IT* | | | | |
| 0 | 36.90% | 25.21% | 9.02% | 56.78% | 56.45% |
| *N*(0,15) | – | –8.53% | –20.36% | 12.68% | 14.29% |
| *N*(–10,15) | – | – | –12.93% | 25.21% | 24.95% |
| *N*(–20,15) | – | – | – | 43.80% | 43.51% |
| *N*(–10,30) | – | – | – | – | –0.20% |
| | Change in *OT* | | | | |
| 0 | 37.54% | 25.70% | 9.36% | 56.38% | 56.16% |
| *N*(0,15) | – | –8.61% | –20.49% | 12.05% | 13.54% |
| *N*(–10,15) | – | – | –13.00% | 24.41% | 24.23% |
| *N*(–20,15) | – | – | – | 43.00% | 42.79% |
| *N*(–10,30) | – | – | – | – | –0.14% |

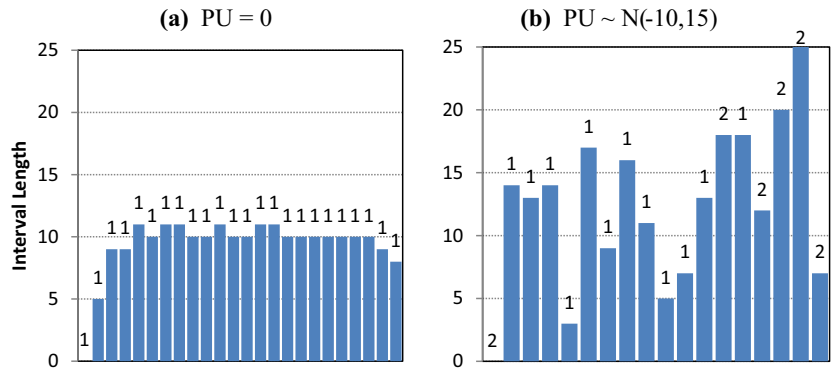
[a]Averaged across $c_{it} = 1, 5, 9$.

(averaging over $c_{di} = 1$, 5, and 9) experiences very little change (0.55% worse for $\mu = -10$ and 3.25% worse for $\mu = -20$). Based on these results, it appears easier to adjust for differences in mean unpunctuality than for increased variability. This suggests that a schedule can be designed to account for significantly early patients but it is much more challenging to account for variability in that earliness. It should be noted that the physician is not penalized when patients' mean unpunctuality changes from 0 to –20. Rather, it is the patients' waiting time that increases while DI and DO decrease, resulting in poorer system performance overall. When the standard deviation is 30 for patient arrivals, all performance measures degrade as expected when compared to a standard deviation of 15: PW by 14%, DI by 20%, DO by 18%, and PDWI by 16%.

## Appointment Scheduling Policies for Patient Unpunctuality

Figure 3 shows the best scheduling policies for some representative scenarios for patient unpunctuality. The height of each bar indicates the length of the appointment interval. For example, the first block has a length of zero indicating that the first appointment is scheduled at the beginning of the session. The number of patients

**Figure 3:** Effect of patient unpunctuality on the scheduling policy.
(a) $PU = 0$. (b) $PU \sim N(-10, 15)$.
Note: Physician lateness = 0, $c_{di} = 9$.

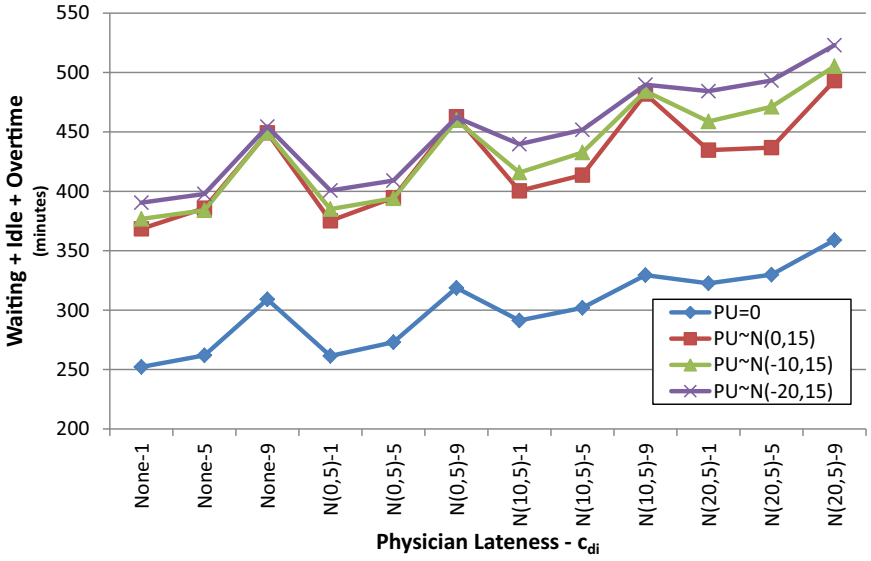



[a] physician lateness=0, $c_{di}$=9

scheduled into that interval is provided above each bar. Each scenario has 24 patients scheduled; multiblock scheduling (e.g., double-booking) results in fewer appointment intervals (i.e., fewer bars in the chart).

Patient unpunctuality has a significant impact on which scheduling policies are best. When patients are punctual (Figure 3a), the results suggest that a plateau dome rule maximizes performance. However, when patients are unpunctual, the schedules are more complex with instances of block scheduling and highly variable intervals (ranging from 3 to 25 minutes). In particular, patients are spread out early in the session (in order to reduce waiting time), whereas increased block scheduling cluster patients later in the session (to avoid excessive overtime). These features are more prominent when patients are earlier on average. We denote this general policy of increasing appointment length and increasing clustering the "increasing interval & clustering rule" (IICR). In general, when patient unpunctuality is present, the system compensates by clustering small groups of patients in short intervals in the first half of the session and larger groups of patients for longer intervals in the second half of the session (Figure 3b). This policy holds for all levels of the mean and standard deviation of patient unpunctuality and we can identify specific tendencies as these factors change. As the standard deviation of unpunctuality increases from 15 to 30, appointment slots are shifted earlier throughout the schedule, compensating for the increased uncertainty. As a result, the last appointment slot is placed significantly earlier (often 15 minutes earlier). In contrast, as mean patient earliness changes from 0 to –20 minutes, appointment slots are shifted later; patients are expected to show up sooner in relation to their appointment time, so they are scheduled later throughout the session. Experiments suggest that it is best to schedule the last appointment approximately eight minutes later when the mean is –20 minutes compared to –10 minutes. Appointment policies for selected scenarios are provided in Appendix A.

**Figure 4:** Effects of patient unpunctuality and physician lateness.



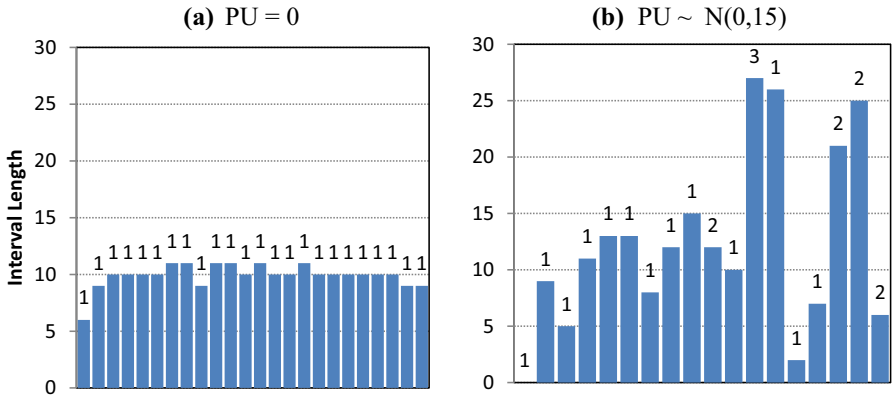

## Impact of other Factors and Performance Measures

### *Physician lateness*

Table 2 and Figure 4 show representative results when physician lateness is added as a factor. As expected, Table 2 shows that as $c_{di}$ increases in value, patient waiting increases, physician idle time and overtime decrease. Averaging over all scenarios, improvements for the doctor are obtained (DI improves by 49.4% and DO by 42.3% as $c_{di}$ increases from 1 to 9) at a cost to patients (PW is 36.1% worse).

The results show that mean physician lateness has a surprisingly large impact on performance. In prior work it has been shown when designing an appointment schedule that an increase in the variability of an input generally has a much larger negative impact than a negative increase in its mean (e.g., Klassen & Rohleder, 1996; Robinson & Chen, 2003). In the presence of unpunctuality, both the mean and variance of physician lateness appear to have a large effect. Adding variability to physician lateness (i.e., increasing the standard deviation from 0 to 5) resulted in deterioration of PWDI of 1.8% (averaging over $c_{di} = 1$, 5, and 9). This involves a reduction in DI (2.2%) and an increase in PW (3.2%) and DO (4.6%). However, as mean lateness increases from 0 to 10 minutes, PWDI deteriorates by 5.2% (PW increases by 8.0%, DI decreases by 4.1%, and DO increases by 22.5%). This is because the physician still has the same amount of work if they show up late, but the appointments are constrained to fit within the scheduling period. The result is that patients are more clustered and overtime increases because all patients must be served. Physician lateness and variability in lateness have a cumulative effect; PW, DO, and overall performance continue to deteriorate whereas DI continues to

**Figure 5:** Best policies for physician lateness. (a) $PU = 0$. (b) $PU \sim N(0, 15)$. Note: Physician lateness $\sim N(10, 5)$, $c_{di} = 9$.




[b] physician lateness ~ N(10,5), $c_{di} = 9$

improve. DO is consistently affected the most, as it is 56.0% worse when physician lateness is $N(20, 5)$ than when it is 0. Thus, the results above suggest that an increase in mean doctor lateness is expected to have at least as much impact as an increase in its variability.

In terms of patient unpunctuality, a change in the mean from –10 to –20 minutes results in a worsening of PWDI by 2.35% for a standard deviation of 15 and 1.54% for a standard deviation of 30. As compared to the case of patient unpunctuality alone, the results are consistent (although less in terms of magnitude) indicating a change in variability has a larger effect on performance. A change in the standard deviation from 15 to 30 results in a worsening of PWDI by 10.28% for a mean of –10 minutes and 9.40% for a mean of –20 minutes. Some representative scheduling policies for physician lateness are shown in Figure 5.

Figure 5a shows that the best scheduling policy with physician lateness when patients are punctual follows the plateau dome rule. In comparison with Figure 3a, the appointment intervals are shorter. In addition, many of the scenarios with physician lateness place the first appointment after time zero, which is logical to reduce patient waiting time. This results in more appointments scheduled later in the session and shorter intervals throughout the session. In Figure 5b, patient unpunctuality is added. Even with a mean unpunctuality of zero, patient unpunctuality has an impact on the best schedule. The results show that a policy that follows the IICR maximizes performance. When combined with physician lateness the variability of patient arrivals results in the first appointment set at time zero even though the physician is late. However, as mean patient earliness increases, the initial appointment is scheduled later than time zero to accommodate physician lateness. Representative schedules with physician lateness can be found in Appendix A and schedules with both physician lateness and patient unpunctuality are given in Appendix B.

### *No-shows and interruptions*

An earlier study by Klassen and Yoogalingam (2009) demonstrated that no-shows did not affect the ideal scheduling policy and the plateau dome rule remained prominent. In this study, no-shows were modeled at two levels: 12.5% and 25% probability of each individual patient not showing up. Although overbooking can be implemented in different ways to mitigate the effects of no-shows (Liu et al., 2010), for this study more patients were scheduled to maintain a constant average utilization level. Although the results show that no-shows have a negative effect on performance, the best overall appointment policy is very similar to those found for prior scenarios with patient unpunctuality and follows the IICR. The difference with no-shows is that patients are scheduled such that the average appointment interval is shorter.

Similar results were found for scenarios that included interruptions. Interruptions degrade performance (e.g., 18.7% worse on average for a 20% interruption rate). However, the best overall scheduling policy with patient unpunctuality is still the IICR.

### *Performance with PWDO, PWDIDO, and higher values for $c_{di}$ and $c_{do}$*

Some of the schedules above (especially for lower levels of $c_{di}$) resulted in a significant amount of overtime for the clinic. Consequently, the performance measures PWDO and PWDIDO were used to determine if the resulting scheduling policies would result in less overtime. Also, in order to focus and fully explore the impact of patient unpunctuality, we have only reported results for $c_{di}$ of 1, 5, and 9. Here we consider the impact of higher weights on the doctor portion of the performance measures to determine if this would impact overtime, patient waiting, idle time, and the appointment policy. In this section we consider higher values for $c_{di}$ and $c_{do}$ of 20 and 50.

The results for PWDO and PWDIDO are similar in both policy (IICR) and performance to those for PWDI (i.e., there is no reduction in overtime). Table 4 has representative results and demonstrates that when idle time is reduced, so is overtime because more patients are seen earlier and fewer are left at the end of the session. Table 4 shows that when PWDO is used, PW is reduced whereas DI and DO are slightly increased (as shown in Klassen & Yoogalingam, 2009). By contrast, when PWDIDO is used, PW is increased whereas DI and DO are lower indicating that DI and DO are closely linked.

The differentiating factor in these scenarios is the cost weighting (e.g., runs with $c_{di} = 5$ are similar to those with $c_{do} = 5$, those with $c_{di} = 20$ are similar to $c_{do} = 20$). Given that these measures produce similar performance and policies, other results are as expected. As $c_{di}$ and $c_{do}$ increase, performance changes in the expected direction in all cases; PW increases and DI and DO both decrease. In terms of scheduling pattern, higher levels of $c_{di}$ and $c_{do}$ cause intervals to be shorter, appointments are bunched more, and the last appointment is scheduled earlier. Note that cases with $c_{di}$ and $c_{do}$ equal to 1 and 50 are less common. For example, a very high value for $c_{di}$ or $c_{do}$ could occur where the physicians have higher priority duties away from their clinics making any idle time or overtime extremely costly. This is reflected in the appointment policies generated which

**Table 4:** Comparison of mean performance with PWDI, PWDO, and PWDIDO.

| $PU\sim(-10,15)$ | $\underline{C_{di}} = 5$ | $\underline{C_{di}} = 9$ | $\underline{C_{di}} = 20$ | $\underline{C_{do}} = 5$ | $\underline{C_{do}} = 9$ | $\underline{C_{do}} = 20$ | $\underline{c_{di}}, \underline{c_{do}} = 5$ | $\underline{c_{di}}, \underline{c_{do}} = 9$ | $\underline{c_{di}}, \underline{c_{do}} = 20$ |
|---|---|---|---|---|---|---|---|---|---|
| *PW* | 315.82 | 408.37 | 555.42 | 315.97 | 382.84 | 460.97 | 423.74 | 503.75 | 611.74 |
| *DI* | 34.53 | 20.52 | 8.68 | 35.00 | 23.05 | 15.41 | 19.08 | 11.90 | 6.41 |
| *DO* | 33.66 | 20.52 | 14.36 | 34.24 | 22.63 | 16.87 | 19.55 | 15.38 | 13.58 |
| PWDI | 488.49 | 593.04 | 729.06 | | | | | | |
| PWDO | | | | 487.19 | 586.49 | 798.30 | | | |
| PWDIDO | | | | | | | 616.87 | 749.31 | 1011.45 |
| $PU\sim(-20,15)$ | | | | | | | | | |
| *PW* | 338.29 | 414.47 | 559.25 | 336.33 | 388.85 | 451.62 | 407.06 | 472.43 | 588.45 |
| *DI* | 30.18 | 19.91 | 8.18 | 30.66 | 22.44 | 15.79 | 20.14 | 13.85 | 6.95 |
| *DO* | 29.35 | 20.05 | 14.05 | 29.85 | 22.13 | 17.10 | 20.23 | 16.22 | 13.64 |
| PWDI | 489.19 | 593.63 | 722.89 | | | | | | |
| PWDO | | | | 485.59 | 588.00 | 793.61 | | | |
| PWDIDO | | | | | | | 608.95 | 743.07 | 1000.31 |

**Table 5:** Results for best scheduling policies.

| Performance Measure | $PU \sim N(-10,15)$ | $PU \sim N(-10,30)$ | $PU \sim N(-20,15)$ | $PU \sim N(-20,30)$ |
|---|---|---|---|---|
| PWDI | | | | |
| *PW* | 568.02 | 579.77 | 575.84 | 642.36 |
| *PW* after appt time | 356.26 | 320.06 | 244.52 | 292.93 |
| *DI* | 8.9 | 13.5 | 8.26 | 9.62 |
| *DO* | 14.28 | 16.5 | 14.19 | 14.77 |
| PWDI (*PW* measured from $t_i$) | | | | |
| *PW* | 580.44 | 695.04 | 627.89 | 729.57 |
| *PW* after appt time | 363.2 | 410.15 | 276.3 | 333.58 |
| *DI* | 7.77 | 7.53 | 6.66 | 6.36 |
| *DO* | 13.53 | 14.07 | 13.69 | 13.53 |

were considerably different than for the middle three levels. For instance, $c_{di} = c_{do} = 1$ results in a large block of patients booked at time 240 and $c_{di} = c_{do} = 50$ results in the last patients being booked much earlier (up to 58 minutes) than the end of the session.

### Performance when Patient Waiting is measured from Appointment Time (*PW*($t_i$)*DI*)

When the objective function includes only waiting after the scheduled appointment start time, not surprisingly the waiting time that is recorded is reduced since any waiting prior to the appointment time is not accounted for. Scenarios were tested using this performance measure including mean patient unpunctuality levels of –10 and –20 minutes and standard deviation values of 15 and 30 for $c_{di}$ and $c_{do}$ levels of 5, 9, and 20. Both total *PW* and *PW*($t_i$) were recorded for the best policies found for the performance measure *PW*($t_i$)*DI*. For comparison, the same was done for the best policies for PWDI. Some representative results are shown in Table 5. Appointments are on average scheduled earlier and (as happens in any appointment system with earlier appointments) idle time and overtime are reduced and total waiting time is increased. It is reasonable that total waiting time increases if the objective function only includes waiting after the appointment. Note, however, that waiting time after the scheduled appointment time also increases (22% on average) because patients are scheduled to show up earlier. The instances of block scheduling are fewer (compared to Figure 3b). However, many of the scheduling policies are still characterized by variable intervals. Thus, measuring waiting only after the appointment time modifies the scheduling policy to favor the server a little more.

## General Rules

It is apparent that patient unpunctuality has a significant and consistent effect on scheduling policy. However, the scheduling rules presented in Figures 3 and 5 and Appendices A and B are not easily implemented by practitioners. It is desirable to develop a set of general rules that are easy to implement, follow the characteristics of the IICR, and could apply to a wide range of scenarios. A number of general rules are presented in Table 6. IICR 1–4 were developed for this study and the dome, plateau dome, and Bailey's rule have performed well in prior studies. The rules are presented as appointment intervals; a zero-length interval means the patient is scheduled at the same time as the patient before (i.e., double booking). Note that this does not represent overbooking because utilization is the same as before. With 24 appointments there are 23 intervals and for these general rules physician lateness is set to zero so the first appointment is always scheduled at time zero. To illustrate, for IICR-1, the first patient is scheduled for time zero, the second 5 minutes into the session, the third 20 minutes in, and so on. The performance (average loss) of each has been measured against the best policies found for scenarios that include the two levels of patient unpunctuality ($\mu = -10$ and –20 minutes), the two performance measures (PWDI and PWDO) and the three middle levels of $c_{di}$ and $c_{do}$ (5, 9, and 20).

IICR-1 is the best performing general rule identified, with only 5.12% loss compared to a range of best rules. It was developed by attempting to mimic the solutions found by the simulation optimization algorithm, with the practical constraints that all appointments had to start on a 5-minute interval and that it should follow a much more consistent pattern. IICR-2 is similar to IICR-1, except the 5-minute interval constraint was not used, and as a result, it was possible to develop a more consistent pattern than IICR-1. IICR-3 was developed for those clinics that do not wish to double-book and want to use the 5-minute intervals; it performs reasonably well by alternating short and long appointments. Similar to IICR-1 and 2, this rule requires quadruple-booking at approximately 225 minutes into the session where four patients are scheduled 15 minutes before the end of the session, and all will show up on average 10 or 20 minutes early. There will be more waiting for these patients, but this rule reduces the probability that the physician will be idle at the end of the session. Based on conversations with family doctors, some prefer double-booking and find it allows them to see patients more efficiently. Multiple or block booking is common in many health care practices because it minimizes the likelihood that the facility (which has a relatively higher cost) will be idle (Rohleder et al., 2011). IICR-4 was developed for clinics that prefer double and triple booking; it is simple to implement. Finally, note that all four rules involve increasing intervals (and an increasing difference between the short and long slots) and larger blocks or clustering as the schedule progresses.

It is not necessarily intuitive that these four rules would perform well. To demonstrate that the IICR rules are better, they were compared to those that have performed well in the literature. Table 4 shows that Bailey's Rule is relatively poor in this environment, performing 20.91% worse. Thus, if waiting before the appointment time is important, Bailey's rule is not a good choice. Comparing it to the other general rules, it is apparent that it will result in increased PW at the

**Table 6:** Appointment intervals for some general rules.

| | | Appointment Interval | | | | | | | | | | | | | | | | | | | | | | | Avg. Loss |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | |
| IICR-1 | Interval length | 5 | 15 | 5 | 20 | 5 | 15 | 5 | 20 | 5 | 15 | | 30 | | | 30 | | | 30 | | 25 | | | | 5.12% |
| | Num of patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 3 | | | 3 | | | 2 | | 4 | | | | |
| IICR-2 | Interval length | 5 | 17 | 5 | 17 | 5 | 17 | 5 | 17 | 5 | 17 | | 29 | | | 29 | | | 29 | | 29 | | | | 5.41% |
| | Num of patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 3 | | | 3 | | | 2 | | 4 | | | | |
| IICR-3 | Interval length | 5 | 15 | 10 | 15 | 10 | 15 | 5 | 15 | 10 | 5 | 10 | 20 | 5 | 5 | 20 | 10 | 5 | 15 | 10 | 20 | | | | 5.14% |
| | Num of patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | | | | |
| IICR-4 | Interval length | 0 | 22 | | 22 | | 22 | | 22 | | 22 | | 29 | | | 29 | | | 29 | | | 29 | | | 5.99% |
| | Num of patients | 1 | 2 | | 2 | | 2 | | 2 | | 2 | | 3 | | | 3 | | | 3 | | | 3 | | | |
| Bailey | Interval length | 0 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 20.91% |
| | Num of patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| Plateau-dome | Interval length | 3 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 10 | 1 | 0 | 6.14% |
| | Num of patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |
| Dome | Interval length | 9 | 9 | 10 | 10 | 10 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 11 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 8 | 0 | 6.70% |
| | Num of patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | |

beginning of the session and increased DI at the end of the session, which will also increase DO. In order to compare the dome and plateau-dome rules fairly, constraints were used to force the algorithm to search for the best solution with a plateau-dome or dome pattern. Both rules performed well, with the plateau dome only approximately 1% worse than the best IICRs. Note that IICRs 1, 2, and 3 are statistically better than the plateau dome (at 95% confidence). For the best performance IICR 1, 2, or 3 can be used. IICR-4 and the plateau dome are good options (not best, but good) depending on clinic preference; if a clinic prefers either a double and triple booking pattern or a more consistent pattern, respectively.

It is also instructive to determine how these rules perform when patient waiting is measured from scheduled appointment time. When using $PW(t_i)$ as a performance measure, the results show there is no rule that performs well across all the environmental factors studied. For example, for the case where mean patient unpunctuality is –20 minutes and the standard deviation is 15, IICR 1–3 perform best for a $c_{it} = 5$, but the dome and plateau dome are best for a $c_{ot} = 20$. In general, IICRs 1–3 perform better for low-cost levels (e.g., $c_{it} = 5$, $c_{ot} = 5$), and the dome and plateau dome perform better for higher cost levels. Overall, the performance of the rules tested resulted in a loss, averaging across the best rules for each scenario and mean unpunctuality levels of –10 and –20, of 12.57%. Accordingly, policies in addition to scheduling (e.g., phone or e-mail reminders) may be warranted to encourage more punctual arrivals.

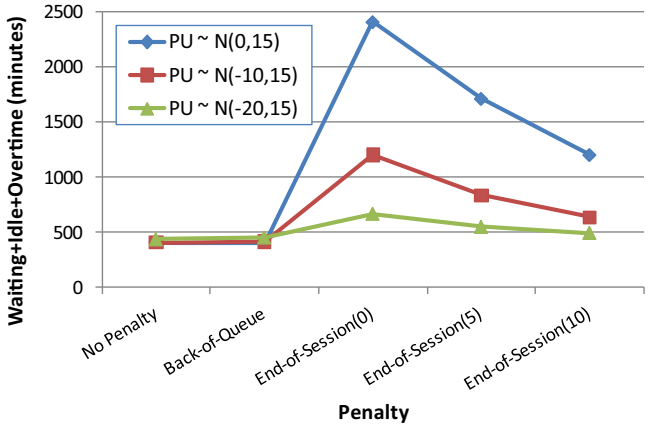
A general rule that accounts for no-shows by overbooking has also been considered and the IICR principles apply, but all appointment intervals are reduced and more slots scheduled in order to maintain utilization levels. For example, for the 24 appointment environment modeled here, a 25% no-show level requires booking 30 appointments. If IICR-4 is used, our experiments suggest that the duration of the double booked slots should be 3 minutes less and the triple booked slots should be 5 minutes less.

## Priority Rules and Late Penalties

Some clinics use priority rules for handling patients that arrive early or late for their scheduled appointment (Powers, 2011). In this study, several rules were considered to determine their impact, if any, on system performance. These rules specify a penalty (or queue discipline for late patients) and a "late allowance" for patients (i.e., how late do they need to be before the penalty is invoked). Late allowances were set at 0, 5, and 10 minutes. Penalties included (i) no penalty, (ii) late patients are sent to the end of the current (existing) queue, (iii) late patients are required to wait until the physician is idle, and (iv) late patients are placed at the end of the session.

These priority rules were tested at varying levels of patient unpunctuality and a full factorial ANOVA was run in order to examine the effects. For brevity, the statistical results are not presented here. However, the main effect of each factor is significant at $\alpha = 0.05$, and based on the sum of squared values, the penalty has the largest effect on performance, followed by patient unpunctuality and the late allowance. Figure 6 summarizes performance of each rule for each level of patient unpunctuality.

**Figure 6:** Comparison of priority rules.




In general, the higher the mean patient earliness and the higher the late allowance, the better the performance. This is because in both cases the penalty is invoked less often. These results suggest that a clinic that is too strict will end up punishing itself and its clients with higher PW, DI, and DO. Performance also declines as the penalty becomes more severe; the end-of-session policy results in the poorest performance. Finally, invoking an end-of-queue policy produces similar results to not having a penalty; DI and DO are the same, with the only difference being a slight increase in PW. However, current session performance is best without any priority rules or penalties. Thus, although using penalty policies *may* encourage patients to show up more punctually in the future, the effect is negative. This research presents the IICR policy which can mitigate the effects of both earliness and lateness and avoid the complexity of a penalty policy.

## DISCUSSION

This article demonstrates the impact of patient unpunctuality on performance and appointment policy design. Because most empirical data collected to date suggests that patients are indeed unpunctual (it is quite challenging for anyone to arrive exactly on time), it is worthwhile to develop policies that take advantage of early client arrivals while minimizing the impact of late arrivals. Policy makers can benefit from considering patient waiting time before appointment time in order to improve clinic performance and reduce congestion in the waiting room which is an indicator of perceived quality of service (Janikaraman et al., 2011). This study provides several insights into appointment scheduling policies in the presence of patient unpunctuality and identifies a number of general principles.

First, as the standard deviation of patient unpunctuality increases, clinics will benefit by reducing interval size throughout the schedule (relative to a schedule with no patient unpunctuality) and compensating for the increased uncertainty. In contrast, as mean unpunctuality increases (i.e., patients show up earlier on

average), clinics can benefit by increasing interval size slightly. Furthermore, if the clinic desires a higher relative weight on the physician's time (i.e., higher levels of $c_{di}$ and $c_{do}$), the clinic should reduce interval size, thereby clustering patients more throughout the session.

Second, when physician lateness is accounted for, appointment intervals are scheduled later in the session and are reduced in length in order to fit all appointments into the scheduling period. This generally results in more patient waiting and clinic overtime, but less physician idle time. The interaction between patient earliness and physician lateness is also important; the more they both increase, the later the initial appointment should be scheduled. Although scheduling the first patient later than the start time of the clinic will likely not be done in practice, this study shows how performance can be improved if a physician is chronically late in starting the session and that this is more important to consider with patient earliness.

Third, in terms of the scheduling policy, it was shown that patient unpunctuality is a factor that makes a difference in terms of which appointment policy is best. Despite some differences in the specific policies generated for various patient unpunctuality scenarios, it was possible to develop general policies for implementation with minimal loss of performance. Prior studies have considered different methodologies, objective functions, service duration distributions, clinic sizes, clinic lengths, cost weightings between physician and patient, end-of-day restrictions, and various levels of patient no-shows, physician interruptions, and physician lateness (e.g., Wang, 1997; Robinson & Chen, 2003; Denton & Gupta, 2003; Kaandorp & Koole, 2007; Klassen & Yoogalingam, 2009). With few exceptions these studies all found the dome or plateau-dome rule to be best over all these factors. This study finds that a policy that includes increasing intervals and increasing clustering (IICR) can improve system performance when the waiting time of the patient before appointment time is considered in the objective function. This policy alternates between long and short intervals and as the session progresses it benefits from an increasing difference between these long and short appointment slots such that eventually it uses double-booking and possibly triple-booking. A few versions of this rule were presented, including a best rule, a rule for clinics that do not wish to double-book, and one for those that prefer to double book. It was also shown that although no-shows, physician lateness, and interruptions negatively impact performance, they do not impact which overall scheduling policy is best; the IICR can be adjusted to the nuances of the clinic in question as explained above. When the waiting time of the patient before their scheduled appointment time is not included the objective function, the IICR policies result in poorer performance, although still perform better than others for lower cost levels. For higher cost levels, the dome and plateau dome perform better in this case.

Future work could delve more deeply into the reasons for patient unpunctuality. This information could facilitate the development of priority rules and/or penalties for mitigating the effects of unpunctuality on the system. In addition, the generalizability of the rules proposed may be investigated across different clinic sizes and session lengths. Another avenue of research is to consider multistage systems, where most patients see a nurse practitioner before seeing the doctor. It is well known that uncertainty in patient waiting time and psychological factors play

a role in the overall service experience (Maister, 1985). Further investigation of these factors and the inclusion of additional phases to the process would be useful in determining if a scheduling rule can be formulated to mitigate the effects of patient unpunctuality.

## REFERENCES

Alexopoulos, C., Goldsman, D., Fontanesi, J., Kopald, D., & Wilson, J. (2008). Modeling Patient Arrivals in Community Clinics. *Omega*, *36*(1), 33–43.

Andradóttir, S. 2006. Simulation optimization with countably infinite feasible regions: efficiency and convergence. *ACM Transactions on Modeling and Computer Simulation*, *16*(4), 357–374.

Babes, M., & Sarma, G. (1991). Out-patient queues at the Ibn-Rochd Health Center. *Journal of the Operational Research Society*, *42*(10), 845–855.

Bailey, N. (1952). A study of queues and appointment systems in hospital out-patient departments with special reference to waiting times. *Journal of the Royal Statistical Society*, *14*(2), 185–199.

Begen, M.A., & Queyranne, M. (2011). Appointment scheduling with discrete random durations. *Mathematics of Operations Research*, *36*(2), 240–257.

Blanco White, M., & Pike, M. (1964). Appointment systems in outpatients' clinics and the effect on patients' unpunctuality. *Medical Care*, *2*(3), 133–145.

Brahimi, M., & Worthington, D. (1991). Queuing models for outpatient appointment systems: a case study. *Journal of the Operational Research Society*, *42*(9), 733–746.

Cayirli, T., & Veral, E. (2003). Outpatient scheduling in health care: a review of literature. *Production and Operations Management*, *12*(4), 519–549.

Cayirli, T., Veral E., & Rosen H. (2006). Designing appointment scheduling systems for ambulatory care services. *Health Care Management Science*, *1*(9), 47–58.

Cayirli, T., Veral, E., & Rosen, H. (2008). Assessment of patient classification in appointment system design. *Production and Operations Management*, *17*(3), 338–353.

Cox, T., Birchall, J., & Wong, H. (1985). Optimising the queuing system for an ear, nose and throat outpatient clinic. *Journal of Applied Statistics*, *12*(2), 113–126.

Denton, B., & Gupta, D. (2003). A Sequential bounding approach for optimal appointment scheduling. *IIE Transactions*, *35*(11), 1003–1016.

Fetter, R., & Thompson, J. (1966). Patients' waiting time and doctors' idle time in the outpatient setting. *Health Services Research*, *1*(1), 66–90.

Fu, M. C. 2002. Optimization for simulation: theory vs. practice. *INFORMS Journal on Computing*, *14*(3), 192–215.

Gupta, D. & Denton, B. (2008). Appointment scheduling in health care: challenges and opportunities. *IIE Transactions*, *40*(9), 800–819.

Janakiraman, N., Meyer, R. & Hoch, S. (2011). The psychology of decisions to abandon waits for service. *Journal of Marketing Research*, *48*, 970–984.

Kaandorp, G., & Koole, G. (2007). Optimal outpatient appointment scheduling. *Health Care Management Science*, *10*(3), 217–229.

Klassen, K., & Rohleder, T. (1996). Scheduling outpatient appointments in a dynamic environment. *Journal of Operations Management*, *14*(2), 83–101.

Klassen, K., & Yoogalingam, R. (2008). An assessment of the interruption level of doctors in outpatient appointment scheduling. *Operations Management Research*, *1*(2), 95–102.

Klassen, K., & Yoogalingam, R. (2009). Improving performance in outpatient appointment services with a simulation optimization approach. *Production and Operations Management*, *18*(4), 447–458.

Klassen, K., & Yoogalingam, R. (2013). Appointment system design with interruptions and physician lateness. *International Journal of Operations and Production Management*, *33*(4), 447–458.

Laganga, L., & Lawrence, S. (2007). Clinic overbooking to improve patient access and increase provider productivity. *Decision Sciences*, *38*(2), 251–276.

Laguna, M. (1997). Optimization of Complex Systems with OptQuest. Graduate School of Business, University of Colorado.

Liu, L., & Liu, X. (1998). Block appointment systems for outpatient clinics with multiple doctors. *Journal of the Operational Research Society*, *49*(12), 1254–1259.

Liu, N., Ziya, S. & Kulkarni, V. (2010). Dynamic scheduling of outpatient appointments under patient no-shows and cancellations. *Manufacturing and Service Operations Management*, *12*(2), 347–364.

Maister, D. (1985). The Psychology of Waiting Lines, accessed March 4, 2013, available at: http://davidmaister.com/articles/the-psychology-of-waiting-lines/.

Powers, M. (2011). Reducing patient wait times: Examine your operations to boost efficiency. ENT Today, October, accessed September 2012, available at: http://www.enttoday.org.

Rising, E., Baron, R., & Averill, B. (1973). A system analysis of a university health service outpatient clinic. *Operations Research*, *21*(5), 1020–1047.

Robinson, L., & Chen, R. (2003). Scheduling doctors' appointments: optimal and empirically-based heuristic policies. *IIE Transactions*, *35*(3), 295–307.

Robinson, L., & Chen, R. (2011). Estimating the implied value of the customer's waiting time. *Manufacturing and Service Operations Management*, *13*(1), 53–57.

Rohleder, T.R., Lewkonia, P., Bischak, D.P., Duffy, P., & Hendijani, R. (2011). Using simulation modeling to improve patient flow at an outpatient orthopedic clinic. *Health Care Management Science*, *14*(2), 135–145.

Tai, G., & Williams, P. (2011). Optimization of scheduling patient appointments in clinics using a novel modelling technique of patient arrival. *Computer Methods and Programs in Biomedicine*, *108*(2), 467–476.

Vanden Bosch, P. M., Dietz, D. C., & Simeoni, J. R. (1999). Scheduling customer arrivals to a stochastic service system. *Naval Research Logistics*, *46*(5), 549–559.

Vissers, J. (1979). Selecting a suitable appointment system in an outpatient setting. *Medical Care*, *12*(12), 1207–1220.

Wang, P. (1997). Optimally scheduling *N* customer arrival times for a single-server system. *Computers & Operations Research*, *24*(8), 703–716.

**Reena Yoogalingam** is an associate professor of operations management at the Goodman School of Business at Brock University in St. Catharines, Ontario, Canada. She holds a PhD in Management Science from York University. Her research has appeared in a number of academic journals including *Production and Operations Management, the International Journal of Operations and Production Management, Journal of the Operational Research Society, Journal of Environmental Management, and Environment and Planning B: Planning and Design*. Her research interests are in the application of simulation and metaheuristic techniques to a variety of management problems in health care and environmental policy planning.

**Kenneth J. Klassen** is a Professor of Operations Management at the Goodman School of Business at Brock University in St Catharines, Ontario, Canada. He holds a PhD in Operations Management from the University of Calgary. His current research focuses primarily on healthcare process improvement and cost reduction. He has presented at over 45 international conferences, taught invited seminars on three continents, and has published in numerous scholarly journals including the *Journal of Operations Management, Production and Operations Management Journal, International Journal of Operations and Production Management, Health Care Management Science,* and *Omega: The International Journal of Management Science*. He has also published case studies for teaching, provided content for a number of textbooks, served on editorial review boards for various journals, and received a number of grants including a national research grant.

## APPENDIX A BEST SCHEDULING POLICIES WITH INDIVIDUAL FACTORS

| Scenario | | | | Appointment Interval | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *DU* | *PU* | *c di* | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| $\mu, \sigma$ | $\mu, \sigma$ | | | | | | | | | | | | | | | | | | | | | | | | | |
| 0,0 | 0,0 | 5 | Interval Length | 7 | 9 | 10 | 11 | 11 | 11 | 11 | 11 | 11 | 12 | 10 | 12 | 12 | 11 | 11 | 11 | 10 | 10 | 11 | 10 | 10 | 9 | 9 |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0,0 | 0,0 | 9 | Interval Length | 5 | 9 | 9 | 11 | 10 | 11 | 11 | 10 | 10 | 11 | 10 | 10 | 11 | 11 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 9 | 8 |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0,0 | 0,15 | 5 | Interval Length | 0 | | 14 | 19 | 14 | 6 | 14 | 18 | 11 | 5 | 7 | 27 | | 4 | 26 | 10 | | 21 | | 31 | | 5 | |
| | | | Number of Patients | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 1 | 2 | | 2 | | 2 | | 2 | |
| 0,0 | 0,15 | 9 | Interval Length | 0 | | 7 | 23 | 12 | 8 | 9 | 17 | 10 | 4 | 9 | 24 | 2 | | 28 | 6 | | 21 | | 29 | | | |
| | | | Number of Patients | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 2 | | 2 | | 4 | | | |
| 0,0 | –10,15 | 5 | Interval Length | 2 | 16 | 9 | 16 | 4 | 6 | 13 | 19 | 10 | 6 | 8 | 17 | 10 | 11 | 11 | 15 | 7 | 15 | 3 | 31 | | | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | | | |
| 0,0 | –10,15 | 9 | Interval Length | 0 | 14 | 13 | 14 | 3 | 17 | 9 | 16 | 11 | 5 | 7 | 13 | 18 | | 18 | 12 | | 20 | | 25 | | 7 | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 2 | | 2 | | 2 | | 2 | |
| 0,0 | –20,15 | 5 | Interval Length | 11 | 19 | 6 | 18 | 3 | 19 | 10 | 17 | | 22 | | 27 | | 6 | 29 | 4 | | 7 | 36 | 6 | | | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 2 | | 2 | | 1 | 1 | 2 | | 1 | 1 | 4 | | | |
| 0,0 | –20,15 | 9 | Interval Length | 10 | 13 | 15 | 7 | 12 | 16 | 8 | 8 | 22 | 4 | | 14 | 21 | 11 | | 11 | 9 | 10 | 23 | 4 | 9 | 3 | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 1 | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| 0,5 | 0,0 | 5 | Interval Length | 9 | 10 | 10 | 10 | 10 | 13 | 11 | 11 | 10 | 12 | 10 | 11 | 12 | 12 | 11 | 10 | 10 | 10 | 11 | 10 | 10 | 9 | 8 |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0,5 | 0,0 | 9 | Interval Length | 7 | 9 | 9 | 11 | 10 | 11 | 11 | 10 | 10 | 11 | 10 | 11 | 10 | 10 | 11 | 10 | 10 | 10 | 10 | 9 | 10 | 9 | 9 |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

*(Continued)*

Continued

| Scenario | | | | Appointment Interval | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *DU* | *PU* | *c di* | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| $\mu, \sigma$ | $\mu, \sigma$ | | | | | | | | | | | | | | | | | | | | | | | | | |
| 10,5 | 0,0 | 5 | Interval Length* | 10 | 10 | 10 | 11 | 10 | 12 | 11 | 10 | 11 | 11 | 10 | 12 | 11 | 11 | 11 | 10 | 10 | 10 | 10 | 10 | 10 | 8 | 4 |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 10,5 | 0,0 | 9 | Interval Length* | 9 | 10 | 10 | 10 | 10 | 11 | 11 | 9 | 11 | 11 | 10 | 11 | 10 | 10 | 11 | 10 | 10 | 10 | 10 | 10 | 10 | 9 | 9 |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 20,5 | 0,0 | 5 | Interval Length* | 11 | 10 | 9 | 11 | 10 | 12 | 11 | 10 | 11 | 11 | 10 | 11 | 11 | 11 | 11 | 10 | 10 | 10 | 10 | 9 | 10 | 5 | 0 |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| 20,5 | 0,0 | 9 | Interval Length* | 10 | 9 | 4 | 16 | 10 | 11 | 11 | 9 | 10 | 11 | 10 | 11 | 10 | 10 | 10 | 10 | 10 | 10 | 10 | 9 | 10 | 8 | 5 |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

## APPENDIX B BEST SCHEDULING POLICIES WITH FACTORS COMBINED

| Scenario | | | | Appointment Interval | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *DU* $\mu, \sigma$ | *PU* $\mu, \sigma$ | *c di* | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| 0,5 | 0,15 | 5 | Interval Length | 0 | 4 | 12 | 20 | 13 | 8 | 12 | 18 | 11 | 5 | 8 | 26 | | 6 | 23 | 11 | | 21 | | 32 | 1 | | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 1 | 2 | | 2 | | 1 | 3 | | |
| 0,5 | 0,15 | 9 | Interval Length | 0 | | 11 | 19 | 13 | 8 | 10 | 17 | 10 | 3 | 10 | 24 | | | 30 | 4 | | 23 | | 25 | | 6 | |
| | | | Number of Patients | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | | | 1 | 2 | | 2 | | 2 | | 2 | |
| 0,5 | –10,15 | 5 | Interval Length | 2 | 18 | 10 | 16 | 13 | 6 | 12 | 21 | 9 | 6 | 8 | 17 | 12 | 4 | 22 | 11 | | 21 | | 32 | | | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 2 | | 4 | | | |
| 0,5 | –10,15 | 9 | Interval Length | 2 | 12 | 13 | 13 | 12 | 9 | 9 | 17 | 11 | 2 | 10 | 24 | 2 | | 25 | 12 | | 17 | | 28 | | 6 | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 2 | | 2 | | 2 | | 2 | |
| 0,5 | –20,15 | 5 | Interval Length* | 11 | 17 | 16 | 11 | 3 | 19 | 9 | 17 | | 22 | | 27 | | 5 | 33 | 2 | | 5 | 36 | 6 | | | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 2 | | 2 | | 1 | 1 | 2 | | 1 | 1 | 4 | | | |
| 0,5 | –20,15 | 9 | Interval Length | 12 | 13 | 15 | 8 | 8 | 23 | 4 | 5 | 25 | 4 | | 15 | 21 | | 11 | 16 | 4 | 10 | 21 | 4 | 11 | 2 | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| 10,5 | 0,15 | 5 | Interval Length* | 5 | 16 | | 20 | 12 | 7 | 11 | 23 | 9 | 5 | 9 | 9 | 26 | 9 | | 14 | 8 | 20 | | 32 | | 1 | |
| | | | Number of Patients | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 1 | 2 | | 2 | | 2 | |
| 10,5 | 0,15 | 9 | Interval Length | 9 | 5 | 11 | 13 | 13 | 8 | 12 | 15 | 12 | | 10 | 27 | | | 26 | 2 | 7 | 21 | | 25 | | 6 | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 3 | | | 1 | 1 | 1 | 2 | | 2 | | 2 | |
| 10,5 | –10,15 | 5 | Interval Length* | 18 | | | 25 | 11 | 8 | 11 | 20 | 11 | 3 | 11 | 9 | 24 | 2 | 11 | 11 | 11 | 16 | | 27 | | | |
| | | | Number of Patients | 3 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 4 | | | |
| 10,5 | –10,15 | 9 | Interval Length* | 12 | 5 | 10 | 14 | 12 | 8 | 11 | 16 | 10 | 6 | 7 | 23 | | 1 | 26 | 10 | | 20 | 2 | 25 | | 5 | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 1 | 2 | | 1 | 1 | 2 | | 2 | |

*Continued*

Continued

| Scenario | | | | Appointment Interval | | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| *DU* $\mu, \sigma$ | *PU* $\mu, \sigma$ | *c di* | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 |
| 10,5 | –20,15 | 5 | Interval Length* | 17 | 3 | 7 | 18 | 4 | 22 | 4 | 15 | 17 | 7 | | 26 | 3 | 2 | 26 | 8 | | 14 | 29 | | | | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 1 | 1 | 1 | 2 | | 1 | 5 | | | | |
| 10,5 | –20,15 | 9 | Interval Length* | 17 | 13 | | 5 | 24 | 7 | 4 | 18 | 21 | | | 16 | 22 | 9 | | 15 | 3 | 10 | 36 | | | | |
| | | | Number of Patients | 1 | 2 | | 1 | 1 | 1 | 1 | 1 | 3 | | | 1 | 1 | 2 | | 1 | 1 | 1 | 5 | | | | |
| 20,5 | 0,15 | 5 | Interval Length* | 14 | | | 26 | 11 | 9 | 13 | 19 | | 12 | 10 | 26 | | | 35 | 2 | | 5 | 42 | | | | |
| | | | Number of Patients | 3 | | | 1 | 1 | 1 | 1 | 2 | | 1 | 1 | 3 | | | 1 | 2 | | 1 | 5 | | | | |
| 20,5 | 0,15 | 9 | Interval Length* | 18 | | 11 | 12 | 6 | 22 | 4 | 14 | 16 | 6 | 1 | 27 | | 3 | 25 | 8 | 3 | 10 | 17 | 6 | 15 | 3 | |
| | | | Number of Patients | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | |
| 20,5 | –10,15 | 5 | Interval Length* | 14 | 9 | 11 | 11 | 9 | 15 | 10 | 15 | 11 | 11 | | 18 | 20 | 1 | 11 | 13 | 7 | 16 | 11 | 11 | | | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 4 | | | |
| 20,5 | –10,15 | 9 | Interval Length* | 14 | 3 | 13 | 14 | 10 | 10 | 11 | 16 | 11 | 4 | 8 | 11 | 13 | 10 | 12 | 17 | 1 | 10 | | 32 | | 4 | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 2 | | 2 | |
| 20,5 | –20,15 | 5 | Interval Length* | 11 | 10 | 6 | 13 | 4 | 25 | | 16 | 13 | 16 | | 14 | 11 | | 27 | | 5 | 39 | | | | | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 2 | | 1 | 1 | 2 | | 1 | 2 | | 2 | | 1 | 6 | | | | | |
| 20,5 | –20,15 | 9 | Interval Length* | 10 | 13 | 10 | 7 | 11 | 15 | 3 | 18 | 16 | 4 | 2 | 13 | 22 | 10 | | 20 | | 6 | 30 | 5 | | | |
| | | | Number of Patients | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | | 2 | | 1 | 1 | 4 | | | |