IIE Transactions
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/uiie20

Scheduling doctors' appointments: optimal and empirically-based heuristic policies

LAWRENCE W. ROBINSON [a] & RACHEL R. CHEN [a]
[a] Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853-6201, USA
E-mail: LWR2@cornell.edu or RC72@cornell.edu
Published online: 29 Oct 2010.

To cite this article: LAWRENCE W. ROBINSON & RACHEL R. CHEN (2003) Scheduling doctors' appointments: optimal and empirically-based heuristic policies, IIE Transactions, 35:3, 295-307, DOI: 10.1080/07408170304367

To link to this article: http://dx.doi.org/10.1080/07408170304367

IIE Transactions (2003) **35**, 295–307

0740-817X/03 $12.00 + .00
DOI: 10.1080/07408170390175512

Scheduling doctors' appointments: optimal and empirically-based heuristic policies

LAWRENCE W. ROBINSON* and RACHEL R. CHEN

Johnson Graduate School of Management, Cornell University, Ithaca, NY 14853-6201, USA
E-mail: LWR2@cornell.edu or RC72@cornell.edu

Received May 2001 and accepted February 2002

Consider the problem facing a doctor (or other service provider) who is setting patient appointment times in the presence of random service times. He or she must balance the patients' waiting times (if the appointments are scheduled too closely together) against the doctor's idle time (if the appointments are spaced too far apart). Although this problem is fairly intractable, this paper uses the structure of the optimal solution as the basis for a simple closed-form heuristic for setting appointment times. Over a wide test bed of problems, this heuristic is shown to perform on average within 2% (and generally within 0.5%) of the optimal policy.

1. Introduction

Everyone who has visited their doctor (or a health clinic) has had plenty of time to observe that the scheduled appointment times can be only marginally related to the time at which they actually see the doctor. Anecdotes about patients billing the doctor for the value of their waiting time periodically surface. The problem of appointment scheduling (setting appointment times that balance the patients' waiting time against the doctor's idle time) is thorny and largely unsolved. Even with the assumption of timely patient arrivals, this problem is still too intractable to be solved optimally, apart from a few special cases (e.g., two patients, phase-type service time distributions). There is also a dearth of generally applicable heuristics which are accessible to practitioners. Almost all of the published heuristics are tested only against other heuristics, and not against the optimal policy. Most do not even recognize that the means of the service times can be eliminated from the formulation, and so can yield arbitrarily poor performance.

In this paper we use the optimal appointment times as the basis for a very simple closed-form heuristic, which can be quickly implemented using only a calculator. Our model is quite general, and can accommodate arbitrary mean service times for every patient, as well as an arbitrary common standard deviation for the set of patients. The standardized service time distribution is derived from an empirical distribution of surgery times. We show that this heuristic generally averages within 0.5% of the cost of the optimal policy, with worst-case performance usually within 20% of the optimal. We also show that our heuristic's performance is fairly robust with regards to distributional misspecification.

More formally, consider setting the appointment times for a sequence of n customers who need to be seen by a service provider over a period of time. Examples of such systems include patients and a medical practitioner, doctors and an operating room, clients and a consulting professional (lawyer or accountant), automobiles and a service center, tractor-trailers and a receiving bay, legal cases and a courtroom, or even students and a professor's office hours. For convenience and clarity, we will use the paradigm of a daily sequence of patients seeing a doctor throughout this paper, although the applicability of this problem is considerably broader. Because the patients' service times are uncertain, the problem of setting appointment times is not a trivial one. Setting them too close together leads to excessive patient waiting times, which is of some slight concern, even for doctors. Setting them too far apart, on the other hand, will cause excessive doctor idle times. The optimal appointment schedule must somehow balance the patients' waiting time against the doctor's. Further, the balance will vary over the course of the day: falling behind in the morning might delay patients throughout the day, while falling behind near closing time might not affect anyone besides the doctor.

Thus the problem is to choose the appointment times to minimize the expected weighted sum of the waiting

*Corresponding author

time for the patients and the idle time of the doctor. We assume that the decision-maker (generally, the doctor) considers patient i's time to be worth some positive fraction α_i of the doctor's time. In other words, we have defined the dimensionality of waiting costs such that the doctor's time has value 1.0. In this paper we considered the range $0.01 \leq \alpha_i \leq 1.0$, although there could certainly be situations where α_i falls outside of this range. For example, the Congressional barber would certainly consider $\alpha_i \gg 1$, whereas financial constraints might force state-run medical clinics to consider $\alpha_i < 0.01$. But in general, we felt that as the decision maker, the doctor (or other service provider) would generally view his or her time to be at least as valuable as the patients' time, so that $\alpha_i \leq 1.0$ is reasonable.

The outline of this paper is as follows. In Section 2, we review some of the most relevant research in this area. In Section 3, we lay out the assumptions behind our problem and formulate the model. We then transform the decision variables from appointment times to job allowances (the time between appointments), and rescale the variables to eliminate the means of the service times from the problem. Finally, we show how Monte Carlo integration can be used to approximate this intractable problem as a stochastic linear program. In Section 4 we develop a wide test bed of problems, and model the service time by fitting the four-parameter generalized lambda distribution to the empirical distribution of Goldman *et al.* (1970). We use the form of the optimal policy as the basis for an atheoretic closed-form heuristic appointment policy. In Section 5 we calculate the increase in cost of the heuristic policy relative to the optimal policy, both in terms of expected value (within 2%, and often 0.5%) and worst-case performance (within 60%, and often 20%). Section 6 describes our overall conclusions and outlines several promising directions of future research.

2. Literature review

The problem of controlling arrivals to queues has been studied since the 1950s, using queuing, dynamic programming, and simulation methodologies. The queuing-based papers by Mercer (1960), Jansson (1966), and Sabria and Daganzo (1989) all study the steady-state behavior of a single server queue with an infinite number of identical customers and constant interarrival times. However, as pointed out by Bailey (1954), these steady-state queuing results are generally inappropriate for appointment systems, which involve only finite number of patients, and do not achieve steady state.

Brahimi and Worthington (1991) use discrete service time distributions to formulate this problem as a Markov chain in discrete time, which allows them to compute the system state probabilities numerically. They use these probabilities to calculate various performance measures for evaluating different appointment systems. They use these measures to approximate the queue characteristics under continuous service time distributions.

Because this problem is analytically intractable, a number of authors have explored various heuristics for assigning patient appointment times. One common scheduling rule is to break up the day into m blocks of equal length, and then assign various numbers of patients to arrive at the beginning of these blocks. Bailey (1952, 1954) and Welch and Bailey (1952) propose 'individual appointment scheduling rules', under which the first k patients are scheduled to arrive at the start of the day, with succeeding patients scheduled at intervals equal to their expected service time μ. Using manual simulation, they conclude that setting $k = 2$ often performs well. More recently, Ho and Lau (1992) use simulation to evaluate the performance of a large number of job allowance heuristics.

Soriano (1966) compares a number of commonly-used heuristics for scheduling appointments in steady state and concludes that a 'two-at-a-time' system, which schedules pairs of patients to arrive simultaneously, is generally best. Blanco White and Pike (1964) allow for unpunctual arrivals, and use simulation to arrive at a similar conclusion. Fries and Marathe (1981) generalize this model to allow for blocks of variable sizes, and use dynamic programming to determine the optimal sizes of each block, assuming that the service times are exponentially distributed. Liao *et al.* (1993) also use dynamic programming. They assume that the service times are Erlang distributed, and allow for dynamic scheduling over the course of the day as service times are revealed.

The above studies assume that the service times are identically distributed for all patients. Charnetski (1984) models heterogeneous patients, where the service time for patient i has a mean of μ_i and a variance of σ_i^2. He evaluates the performance of a heuristic which sets a job allowance of $\mu_i + z\sigma_i$ for each patient, using simulation in his search for the optimal value of the safety factor z.

The shortcoming of all of these heuristics is that this safety factor z is constant. (Ho and Lau (1992) do allow the safety factor z to jump from one value to a second.) These heuristics will not generally be optimal; in particular, Wang (1993) showed that the optimal job allowance times exhibit a 'dome' pattern, rather than being constant, when the service times are exponentially distributed.

More recently, the literature has focused on the optimal scheduling of a finite number of customers. Weiss (1990) shows that when there are only two patients, the problem is analogous to a simple newsvendor problem. Pegden and Rosenshine (1990) consider the case in which the patient service times are exponentially distributed, and derive the optimal job allowances for two and three patients. Wang (1993) generalizes this problem to an arbitrary number of patients, and shows that the customer waiting time will have a phase-type distribution, which

allows him to find the optimal job allowances by solving a set of nonlinear equations. As was previously mentioned, he shows that for a fixed number of identical patients with exponentially-distributed service times, the optimal job allowances are 'dome' shaped, with higher job allowances for patients in the middle of the day. Wang (1997) determines the optimal appointment times when the service time can be approximated by a phase-type probability distribution. Vanden Bosch and Dietz (2001) propose methods for scheduling appointments when customer arrivals are constrained to a discrete lattice of times.

Robinson *et al.* (1996) solve the three-patient problem with general service time distributions. However, their analytical approach for characterizing the optimal job allowances is intractable for more patients. They then propose a stochastic programming approach, which combines Monte Carlo integration with linear programming, which they use to solve the problem for up to 16 identical patients with normally-distributed service times. Their results also show the dome pattern for patient allowance times. Denton and Gupta (2001) present a stochastic linear programming model for determining the optimal appointment schedule when the patient service times follow general discrete distributions. They also develop an algorithm that generates converging upper and lower bounds when the patient service times follow general continuous distributions.

An important practical issue in all of these models is how to determine the appropriate distribution to use in modeling the stochastic patient service time. For reasons of tractability, Jansson (1966), Soriano (1966), Pegden and Rosenshine (1990), and Wang (1993) all assume that the service times are exponentially distributed. Robinson *et al.* (1996) assume that the service times are normally distributed. In an empirical study, O'Keefe (1985) finds that patient service times often have a coefficient of variation $c_v \doteq \sigma/\mu$ between 0.58 and 0.70, but demonstrate a variety of skewness and kurtosis, and so cannot always be characterized by a single simple two-parameter probability distribution. Ho and Lau (1992) use the four-parameter distribution developed by Schmeiser and Deutsch (1977), and observe that the ranking of the performance of their heuristic rules is not affected by the skewness and kurtosis of the patient service time distribution. Although they conclude that only the first two moments of duration distributions are important in determining job allowances, their conclusion is based on the relative performance of a small set of heuristic rules, rather than on the optimal scheduling policy.

Welch (1964), Goldman *et al.* (1970), and Brahimi and Worthington (1991) all provide empirically-derived histograms of the patient service times. All three histograms display the same general form: unimodal and right-skewed. In our numerical study, we will model patient service times as following the Generalized Lambda Distribution (GLD) developed by Ramberg and Schmeiser (1972, 1974) which matches the histogram of Goldman *et al.* (1970). This histogram reflects 1000 observations, as compared to 151 observations for Welch (1964), and 114 observations for Brahimi and Worthington (1991).

3. Model assumptions and development

3.1. *Model assumptions*

The first and strongest assumption of our model is that the patients will arrive promptly at the established appointment times. This is a standard assumption in the scheduling literature, and holds in many important applications; e.g., hospital operating rooms and non-medical professional services (lawyers and accountants). Admittedly, this assumption is less than completely representative of other real systems, particularly (in our personal experience) of doctors' offices. However, we believe that patient lateness in these other settings is largely in response to past experiences of poor scheduling and long wait times. So if a better scheduling system were successful in reducing expected waiting times, patients would start to make more of an effort to arrive on time, as they currently do for other service providers, and this assumption would become more realistic. As a particular example, Schafer (1986) describes the steps he took in his pediatric practice both to ensure prompt arrivals and to maintain his patient schedule. He reports that 99% of his patients are seen at their scheduled appointment time. More generally, Hall (1991) describes a number of techniques for changing both the arrival and service processes in order to reduce delays.

Note that in some situations it may be the case that some patients arrive early, but do not mind waiting prior to their scheduled appointment times. Allowing for early arrivals with no waiting cost is very close to, but not quite the same as, our assumption of prompt patient arrivals, since it would be possible for an early patient to start service before his or her appointment time whenever the doctor is available earlier.

Finally, relaxing this assumption would raise a number of thorny issues in addition to the game-theoretic patient arrival times discussed above. For example, with uncertain arrival times the sequence in which patients are seen by the doctor may differ from the sequence of their appointment times. Additionally, variable arrival times are likely to require a two-tier patient waiting cost, where waiting prior to the appointment time is perceived to be less onerous than waiting beyond it. These issues are substantial enough to warrant their own paper, which is currently under development.

A second assumption is that the sequence of n patients has already been specified. A reasonable rule-of-thumb for sequencing might be to do so in order of increasing variability of their service times, in order to minimize the

negative consequences of uncertainty. However, the optimality of this appealing heuristic has proved difficult to establish. Wang (1993) showed this sequencing rule to be optimal if the service times were exponentially distributed. Robinson *et al.* (1996) showed that this rule is optimal for $n = 3$ patients facing a common distributional form (scaled by both mean and standard deviation). The performance of this (or any other) sequencing rule in more general cases is still an open issue.

Finally, we assume that the doctor is punctual, and has no other responsibilities over the course of the day beyond seeing these n patients. An alternative formulation might consider additional doctors' responsibilities to include a set of 'postponable tasks' (e.g., returning phone calls, arguing with HMO's, etc.), which can occupy idle time and which must in any case be completed by the end of the day.

3.2. *Model development*

For each patient i, we assume that the value of his or her time, relative to the doctor's, is defined by some multiple α_i. Further, we assume that the service time for patient i, ξ_i, is stochastic with a mean of μ_i and a standard deviation of σ_i. We assume that these service times are mutually independent. We can eliminate from consideration any patient $i < n$ with a deterministic service time by 'collapsing' him or her into the following patient in a preprocessing stage as:

$$\mu_{i+1} \leftarrow \mu_i + \mu_{i+1}, \qquad \alpha_{i+1} \leftarrow \alpha_i + \alpha_{i+1}.$$

Although these assumptions are sufficient for the following problem formulation, we would like to identify at this point the additional restrictions that will be imposed later in the paper for the numerical testing. First, we will assume that while the mean service time μ_i can differ from one patient to another, the distribution of the service time about these differing μ_i will be the same for the first $n-1$ patients. In particular, this implies that the first $n-1$ service times are homoskedastic, with a common standard deviation of σ. Note the service time of the final patient, ξ_n, will not affect anyone's waiting time, and so can follow an arbitrary distribution. Second, we will assume that the cost of waiting for the last $n-1$ patients will all be identical, at some common value α. Because the first patient will never wait, the value of his or her time (α_1) can be arbitrary.

For each patient $i = 1, \ldots, n$, define

$A_i \doteq$ appointment (arrival) time for patient i ($A_1 \equiv 0$),

$S_i \doteq$ actual starting time of service for patient i,

$\quad = \max\{A_i, S_{i-1} + \xi_{i-1}\} \quad i = 2, \ldots, n,$

$S_1 = A_1 = 0.$

In other words, patient i will arrive at the appointment time A_i and start service at time S_i, which is the earliest time at which both the patient is available (A_i) and the doctor has finished the previous patient ($S_{i-1} + \xi_{i-1}$). Patient i will finish service at time $S_i + \xi_i$. The waiting time for patient i is simply $S_i - A_i$. The doctor's total waiting time is equal to the sum of the times between finishing service for one patient and starting service for the next,

$$\sum_{i=1}^{n-1}[S_{i+1} - (S_i + \xi_i)] = S_n - \sum_{i=1}^{n-1}\xi_i,$$

since $S_1 = 0$. Equivalently, the doctor's idle time can be calculated as the time the doctor starts examining the final patient (S_n), minus the total time that he or she has spent examining the first $n-1$ patients. The doctor's cost of waiting has been normalized to 1.0. Defining $\mathbf{A} = \{A_2, \ldots, A_n\}$ and $\boldsymbol{\xi} = \{\xi_1, \ldots, \xi_{n-1}\}$ allows the problem to be written as

$$\text{(P1)} \quad \min_{\mathbf{A}}\left\{E_{\boldsymbol{\xi}}\left[\left(S_n - \sum_{i=1}^{n-1}\xi_i\right) + \sum_{i=2}^{n}\alpha_i(S_i - A_i)\right]\right\}, \quad (1)$$

subject to

$$S_i = \max\{A_i, S_{i-1} + \xi_{i-1}\} \quad i = 2, \ldots, n, \quad (2)$$

$$S_1 = 0.$$

The formulation becomes much more tractable if we redefine the decision variables. Define $X_i \doteq A_{i+1} - A_i$ to be the job allowance time for patient i; i.e., it is the time allotted between patients i and $i+1$. We can always recover A_j from $\mathbf{X} = \{X_1, \ldots, X_{n-1}\}$ as $A_j = \sum_{i=1}^{j-1} X_i$.

Also, by explicitly defining $W_i \doteq S_i - A_i$ to be the waiting time for patient i, we can rewrite Equation (2) as

$$\begin{aligned} S_i &= \max\{A_i, S_{i-1} + \xi_{i-1}\}, \\ S_i - A_i &= \max\{0, (A_{i-1} + W_{i-1}) + \xi_{i-1} - A_i\}, \\ W_i &= (W_{i-1} + \xi_{i-1} - X_{i-1})^+. \end{aligned} \quad (3)$$

Noting that

$$S_n = A_n + W_n = \sum_{i=1}^{n-1} X_i + W_n$$

allows (P1) to be rewritten as

$$\text{(P2)} \quad \min_{\mathbf{X}}\left\{E_{\boldsymbol{\xi}}\left[\sum_{i=1}^{n-1}(X_i - \xi_i) + W_n + \sum_{i=2}^{n}\alpha_i W_i\right]\right\}, \quad (4)$$

subject to

$$W_i = (W_{i-1} + \xi_{i-1} - X_{i-1})^+ \quad i = 2, \ldots, n,$$

$$W_1 = 0.$$

At this point we invoke our assumption of homoskedasticity in the service times (with σ denoting this common standard deviation), and rescale both the decision variables and random variables as:

$$x_i \doteq (X_i - \mu_i)/\sigma,$$
$$\zeta_i \doteq (\xi_i - \mu_i)/\sigma, \text{ and}$$
$$w_i \doteq W_i/\sigma;$$

this allows us to factor out the σ from the model. For notational convenience, we redefine $\alpha_n \leftarrow \alpha_n + 1$. With $\mathbf{x} \doteq \{x_1, \ldots, x_{n-1}\}$, we can substitute these redefinitions into (P2) to yield

$$\text{(P3)} \quad \sigma \cdot \min_{\mathbf{x}} \left\{ \sum_{i=1}^{n-1} x_i + \sum_{i=2}^{n} \alpha_i E_\zeta[w_i] \right\}, \tag{5}$$
subject to
$$w_i = (w_{i-1} + \zeta_{i-1} - x_{i-1})^+ \quad i = 2, \ldots, n,$$
$$w_1 = 0,$$

since $E[\zeta_i] = 0$ by definition. Note that the objective function is proportional to σ, and that (P3) is independent of all of the following:

- The expected service times $\mu_1, \ldots, \mu_n$.
- The common standard deviation of the service times, σ.
- The probability distribution of the service time for the last patient, ξ_n.
- The value of time for the first patient, α_1.

The formulation depends only on the transformed random variables $\{\zeta_i\}$, which are i.i.d. with a mean of zero and a standard deviation of one. In essence, we are defining the unit of time such that $\sigma = 1$. The final service time ξ_n will not affect the waiting times for either the doctor or any other patients. Because the first patient will never wait, his or her waiting cost α_1 is irrelevant.

3.3. *Solution methodology*

Generally speaking, the definition of the waiting times as the positive part of another function makes this problem fairly intractable to solve exactly, outside of a couple of special cases. Weiss (1990) shows that for $n = 2$, this problem can be formulated as a newsvendor model. Wang (1993) is able to find the optimal solution for exponentially-distributed service times. More generally, Denton and Gupta (2001) 'discretize' the continuous service time distributions and apply Bender's decomposition to the resultant stochastic linear program to find optimal and heuristic solutions for up to $n = 10$ patients.

An alternative solution methodology is Monte Carlo integration, which replaces the true distribution of $\boldsymbol{\zeta}$ with an approximate discrete distribution consisting of a sample of K randomly generated points $\{\boldsymbol{\zeta}^k\}$. The well-known advantage of Monte Carlo integration is that the goodness of the resulting approximate solution depends on the size of the sample (K) and not on the dimensionality of the underlying problem ($n - 1$); Hammersley and Handscomb (1964), Halton (1970), and Fishman (1996) all provide good introductions to this methodology. In this case, the expected waiting time will be replaced by the average waiting time across the K sampled points. Dropping the extracted σ term, the optimization problem becomes

$$\text{(P4)} \quad \min_{\mathbf{x}} \left\{ \sum_{i=1}^{n-1} x_i + \sum_{i=2}^{n} \alpha_i \left[\frac{1}{K} \sum_{k=1}^{K} w_i^k \right] \right\}, \tag{6}$$
subject to
$$w_i^k = \left(w_{i-1}^k + \zeta_{i-1}^k - x_{i-1} \right)^+ \quad i = 2, \ldots, n, \quad k = 1, \ldots, K, \tag{7}$$
$$w_1^k = 0 \quad k = 1, \ldots, K,$$

where w_i^k is the waiting time for patient i under service time realization $\boldsymbol{\zeta}^k$.

Robinson *et al.* (1996) show that this problem could be formulated as a linear program, by replacing Equation (7) with the pair of inequalities

$$w_i^k \geq w_{i-1}^k + \zeta_{i-1}^k - x_{i-1} \quad i = 2, \ldots, n, \quad k = 1, \ldots, K, \tag{8}$$
$$w_i^k \geq 0 \quad i = 2, \ldots, n, \quad k = 1, \ldots, K, \tag{9}$$

for $k = 1, \ldots, K$. Because the optimization routine will drive the w_i^k terms to their lowest feasible level, each pair of constraints will correspond to Equation (7) above. The linear program is quite large, with $K(n-1)$ constraints, $K(n-1)$ non-negative decision variables, and $n - 1$ unrestricted-in-sign decision variables. However, the fact that it can be represented as a linear program means that it is convex and quite well behaved; a greedy search in $n - 1$ dimensions will arrive at the optimal $\mathbf{x}$. In this paper we used a conjugate gradient search, based (for convenience) on the left-hand partial derivatives

$$\frac{\partial w_\ell^k}{\partial x_i} = \left\{ \begin{array}{ll} -\alpha_\ell & \text{if } \ell > i \text{ and } w_j^k > 0 \ \forall j \in [i+1, \ell] \\ 0 & \text{otherwise} \end{array} \right\}. \tag{10}$$

Intuitively, the choice of x_i will never affect the waiting time of earlier patients. A later patient will only be affected if he or she, and all intervening patients, had to wait; in this situation a marginal increase in the job allowance x_i will decrease the waiting times for all these delayed patients. Note that this partial derivative will be discontinuous whenever $x_{j-1} = w_{j-1}^k + \zeta_{j-1}^k$ for any j between $i + 1$ and ℓ, in which case the right-hand partial derivative would be zero. The right-hand partial derivative cannot be determined from the values of w_j^k, but instead can be defined by replacing the condition $w_j^k > 0$ in Equation (9) with the more-cumbersome $x_{j-1} \leq w_{j-1}^k + \zeta_{j-1}^k$. Define $\ell^k(i)$ to be the index of the last of these delayed patients

$$\ell^k(i) \doteq \max \left\{ \ell > i \mid w_j^k > 0 \quad \forall j \in [i+1, \ \ell] \right\}. \tag{11}$$

Then the partial derivatives of the objective function can be concisely represented as

$$\frac{\partial}{\partial x_i} = 1 - \frac{1}{K} \sum_{k|w_{i+1}^k > 0} \left(\sum_{j=i+1}^{\ell^k(i)} \alpha_j \right). \tag{12}$$

Define $\mathbf{x}^* = \{x_1^*, \ldots, x_{n-1}^*\}$ to be the optimal solution to (P4). In the discussion of the heuristic policies in the following section, this policy will be referred to as the 'optimal' policy, although the reader should keep in mind that (P4) is only an approximation to the original (P3). But because the Monte Carlo integrations used in this research were particularly thorough (employing Cornell's supercomputer to calculate $\mathbf{x}^*$ as the average across 1000 replication of (P4), each of which used $K = 10\,000$ randomly-generated vectors of service times), this distinction will be trivial.

4. Development of the heuristic policy

While Monte Carlo integration can certainly be used to solve this problem, its high computational requirement reduces its applicability in many practical situations. The purpose of this section is to develop a simple and very tractable heuristic for determining the job allowances. Towards this end, we develop a test bed of the parameter values of interest, investigate and characterize the form of the optimal policy over this test bed, derive a simpler approximate form for this optimal policy, and then finally develop a tractable approximation for the policy parameters within this approximate policy.

4.1. *Problem test bed*

Our first step in developing a heuristic policy was to define an appropriate test bed of problems, and calculate the optimal policy for each of those problems. We considered n identical patients, for 10 different values of the number of patients n between three and 16, and 21 different values for the (common) value of patient's waiting time α between 0.01 and 1.0; the full range of values is given in Table 1. We choose 16 as the test bed's upper bound for n because this would allow for half-hour appointments over the course of an 8-hour day, or 15-minute appointments if the morning and afternoon patients were 'decoupled' by a lunch break halfway through the day.

The second step was to decide upon an appropriate continuous probability distribution to model the service times. *A priori*, we felt that in many service settings, and certainly for doctor visits, the distribution of service times would be positively skewed. For lack of a clearly superior alternative, we decided to base our distribution on the empirical study of Goldman *et al.* (1970), who provide a histogram for 1000 sample ratios of actual/forecasted surgery times, which is shown in Fig. 1. Because the means and (common) standard deviation do not affect our optimization (P4), we focus on the third and fourth moments. This histogram shows a moderate degree of skewness ($\alpha_3 = 0.646$), with Pearson kurtosis somewhat greater than a normal distribution ($\alpha_4 = 3.678$, versus 3.0), as identified in Table 2. Although we could have sampled service times directly from the original histogram, this would have resulted in only 17 distinct observations. We instead represented his findings through the four-parameter Generalized Lambda Distribution (GLD) developed by Ramberg and Schmeiser (1972, 1974) and described in detail by Karian and Dudewicz (2000), which is both appealing in shape and quite versatile in matching empirical data. With $f(x|\boldsymbol{\lambda})$ and $F(x|\boldsymbol{\lambda})$ defining the p.d.f. and c.d.f. of the service time respectively, the GLD is defined by four parameters $\boldsymbol{\lambda} = \{\lambda_1, \ldots, \lambda_4\}$ in terms of the inverse c.d.f. as

$$x = F^{-1}(p|\boldsymbol{\lambda}) \doteq \lambda_1 + \frac{p^{\lambda_3} - (1-p)^{\lambda_4}}{\lambda_2}, \tag{13}$$

Table 1. Test bed values

Number of patients (n)	*Value of patient's time (α)*	
3	0.01	0.1
4	0.0125	0.125
5	0.015	0.15
6	0.02	0.2
7	0.025	0.25
8	0.03	0.3
10	0.04	0.4
12	0.05	0.5
14	0.065	0.65
16	0.08	0.8
		1.0

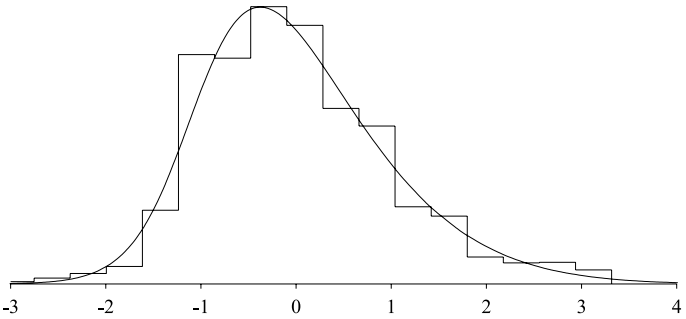

Fig. 1. Standardized distribution of patient service times, after Goldman *et al.* (1970).

Table 2. Skewness and kurtosis for various data sets

Data source	α_3	α_4
Goldman *et al.* (1970)	0.646	3.678
Welch (1964)	0.747	2.958
Brahimi and Worthington (1991)	1.292	4.512
Normal	0.000	3.000

with

$$f(x|\boldsymbol{\lambda}) \doteq 1 \Big/ \left.\frac{\mathrm{d}F^{-1}(p|\boldsymbol{\lambda})}{\mathrm{d}p}\right|_{p=F(x|\boldsymbol{\lambda})} = \frac{\lambda_2}{\lambda_3 F(x|\boldsymbol{\lambda})^{\lambda_3-1} + \lambda_4[1-F(x|\boldsymbol{\lambda})]^{\lambda_4-1}} . \quad (14)$$

We wanted to choose $\boldsymbol{\lambda}$ to maximize the likelihood of observing the empirical data of Goldman *et al.* (1970). Representing their histogram as a set $\{(x_i, \pi_i)\}$, where $\pi_i = \text{Prob}\{X = x_i\}$, then with $N = 1000$ observations, this problem can be represented as

$$\max_{\boldsymbol{\lambda}} \left\{ \prod_i f(x_i|\boldsymbol{\lambda})^{\pi_i N} \right\}. \quad (15)$$

For tractability, we will instead maximize the log of this likelihood function. Substituting in Equation (13) and factoring out the irrelevant N yields the equivalent problem

$$\min_{\boldsymbol{\lambda}} \left\{ \sum_i \pi_i \ln\left[\lambda_3 F(x_i|\boldsymbol{\lambda})^{\lambda_3-1} + \lambda_4[1-F(x_i|\boldsymbol{\lambda})]^{\lambda_4-1} \right] - \ln(\lambda_2) \right\}; \quad (16)$$

the optimal value of $\boldsymbol{\lambda}$ is given in the first row of Table 3, after standardizing to $\mu = 0$ and $\sigma = 1$. Figure 1 shows that this fitted GLD provides an excellent match to the data of Goldman *et al.* (1970). Although there is no reason to believe that this distribution is universally applicable to all service settings, it is at least empirically based. The sensitivity of the heuristic's performance to the service time distribution will be examined in Section 5.

4.2. *Approximating the form of the optimal policy*

The next step of the heuristic development was to investigate the form of the optimal policy, in the hope that a simpler approximate policy form might be able to adequately capture most of the structure of the optimal policy. Each of the 210 test problems was formulated as (P4) with $K = 10\,000$ Monte Carlo generated samples of the service times, and was solved using a conjugate gradient search 1000 times. Define $\mathbf{x}^* = \{x_1^*, \ldots, x_{n-1}^*\}$ to be the average 'optimal' solution, over these 1000 replications. Again, to the degree that (P4) is only an approximation of (P3), this solution will only be an approximation to the true optimal solution.

Two representative cross sections of the 210 solutions are given in Figs. 2 and 3. A number of general observations can be made about the form of these optimal solutions:

- The job allowances follow a 'dome' pattern for all n and α, with more time allotted to patients in the middle of the day. Wang (1993) found this same result for exponentially-distributed service times.
- The first job allowance, x_1^*, is always much lower than the other x_i^*'s, and varies only slightly with n, for α fixed.
- The final job allowance, x_{n-1}^*, is also somewhat lower than the other x_i^*'s.
- The intermediate job allowances, $x_2^*, \ldots, x_{n-2}^*$, are all about the same.
- Unsurprisingly, all of the x_i^*'s increase with α, for n fixed.

This suggests that an approximate policy which uses a single job allowance for all intermediate patients might perform almost as well as the optimal solution. The next step was to investigate, on an *ad hoc* basis for $n = 8$, just how complicated of a policy was required in order to adequately capture the structure of the optimal policy. The four simple policies that were examined are outlined in Table 4. The zero-parameter model sets all job allowance times equal to zero. Two one-parameter models were examined. The first sets all $n-1$ job allowances to the same value, which is equivalent to the heuristic proposed by Charnetski (1984), although his is based on truncated normal distributions for up to 20 patients per day. The second model solves for the first job allowance $\tilde{x}_1$, and sets the remaining $n-2$ job allowances equal to zero. This generalizes the heuristic proposed by Bailey (1952), who proposed that the first two patients be assigned the same appointment time ($A_2 = A_1 = 0$, or $x_1 = -\mu_1/\sigma$), with the others offset by their expected service time ($A_{i+1} = A_i + \mu_i$, or $x_i = 0$, for $i \geq 2$). Note that both one-parameter models are generalizations of the zero-parameter model, and so will always outperform it. The two-parameter model is a combination of the two one-parameter models: It solves for the first job allowance $\tilde{x}_1$, and for single job allowance $\tilde{x}_2$ for each of the remaining $n-2$ patients.

Table 3. Generalized Lambda Distribution parameters ($\mu = 0$, $\sigma = 1$)

Data source	λ_1	λ_2	λ_3	λ_4
Goldman *et al.* (1970)	−0.504 073	0.122 036	0.041 722	0.113 048
Welch (1964)	−0.963 222	0.206 155	0.032 474	0.298 743
Brahimi and Worthington (1991)	−0.831 499	0.049 985	0.006 313	0.050 239
Normal	0.000 000	0.197 451	0.134 912	0.134 912

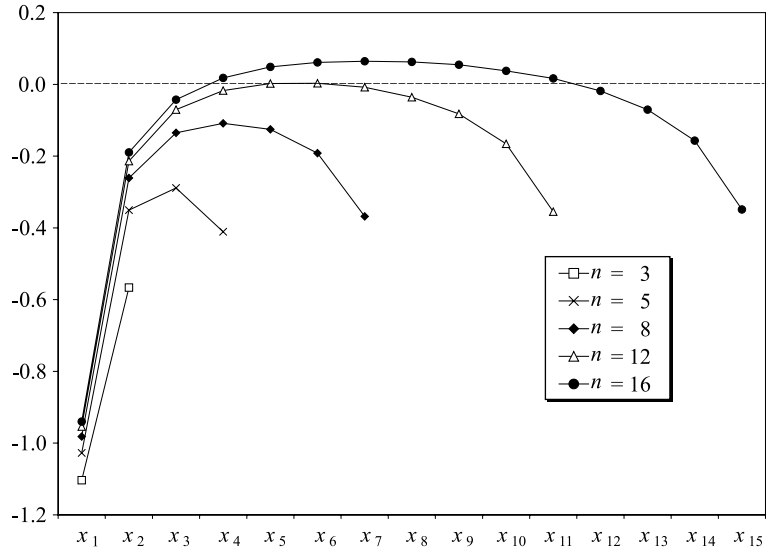

Fig. 2. Representative optimal job allowance times ($\alpha = 0.1$).

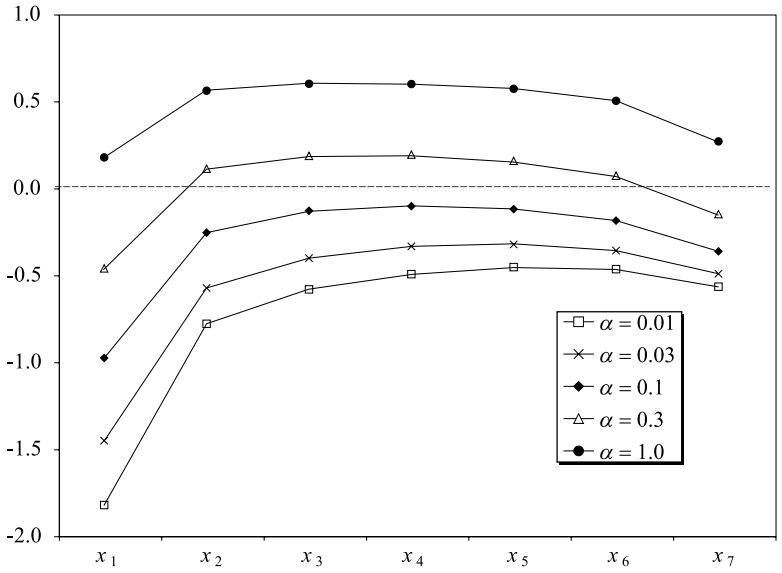

Fig. 3. Representative optimal job allowance times ($n = 8$).

Table 4. Heuristic policies examined

Number of parameters	x_1	$x_2 = x_3 = \cdots = x_{n-1}$
0	0	0
1	$\tilde{x}_1$	$\tilde{x}_1$
1	$\tilde{x}_1$	0
2	$\tilde{x}_1$	$\tilde{x}_2$

The same conjugate gradient search methodology was again used to solve these one- and two-parameter heuristic policies; the average percentage increase in cost above the optimal is shown in Fig. 4 for $n = 8$, as a function of α. It is clear that while both the zero- and one-parameter policies can result in expected costs of more than 20% above optimal, the two-parameter policy consistently performs within 1% of optimal. Similar results were found for other values of n. Because of the strong performance of this two-parameter policy, more complicated policies were not investigated.

4.3. *Approximating the parameters of the two-parameter heuristic policy*

With the form of the heuristic policy approximated by two parameters ($\tilde{x}_1$ for the first job allowance, and $\tilde{x}_2$ for each of the remaining job allowances), the next step was to find the optimal values for these two parameters,

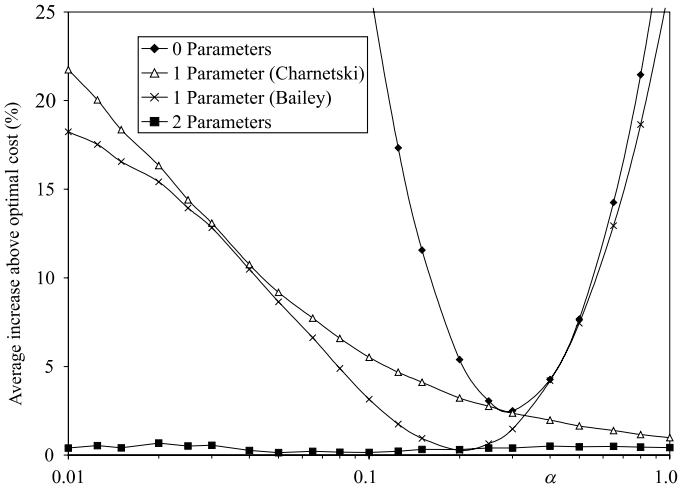

Fig. 4. Simple heuristic performance: average percent increase above optimal cost ($n = 8$).

within this approximate policy, for the 210 test problems. Again, each of these test problems was solved 1000 times, each using $K = 10\,000$ Monte Carlo generated observations of the service times. The final step was to construct two concise atheoretic functions of n and α which approximate the optimal values of these two parameters across the test bed. An afternoon of unstructured experimentation resulted in the following two closed-form atheoretic functions, which approximate the optimal job allowances for the heuristic policy.

$$\hat{x}_1(n,\alpha) \doteq a + b\ln(\alpha), \tag{17}$$

$$\hat{x}_2(n,\alpha) \doteq c + (\alpha^d - c)(n^{-e} + 1), \tag{18}$$

with parameters

$$\begin{aligned} a &= 0.111\ 878,\\ b &= 0.473\ 760,\\ c &= 2.221\ 271,\\ d &= 0.301\ 939,\\ e &= 0.444\ 411. \end{aligned} \tag{19}$$

Note that $\hat{x}_1$ depends on α but not n. The approximate and optimal parameter values are graphed in Figs. 5 and 6 as functions of α, for representative values of n. These figures show that these atheoretic functions do a reasonably good job of fitting the optimal parameter values. For both $\hat{x}_1$ and $\hat{x}_2$, the fit is the worst for $n = 3$, especially for extreme values of α. Table 5 summarizes the sequence of approximations in both policy and parameter values which were taken.

5. Numerical results

In this section we examine the performance of the heuristic policy $\hat{\mathbf{x}}$, relative to the optimal policy $\mathbf{x}^*$. We present both average and worst-case results for the relative

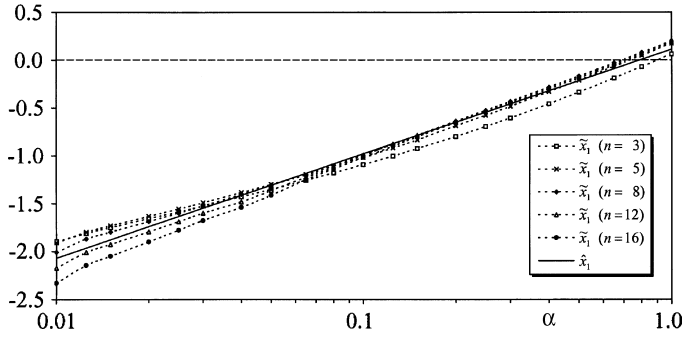

Fig. 5. Atheoretic $\hat{x}_1$ versus optimal $\tilde{x}_1$ parameter values within the heuristic policy.

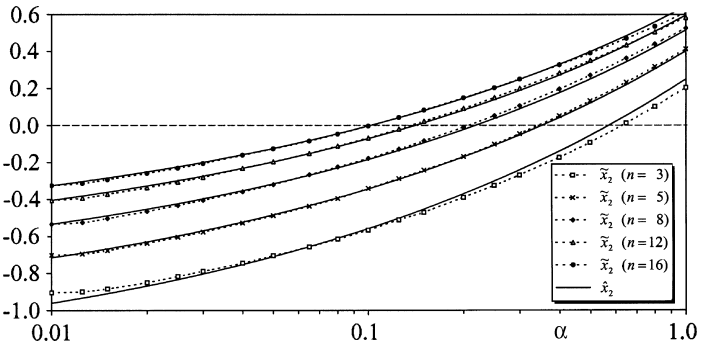

Fig. 6. Atheoretic $\hat{x}_2$ versus optimal $\tilde{x}_2$ parameter values within the heuristic policy.

Table 5. Summary of policies examined

Policy type	*Number of parameters*	*Parameter values*	*Job allowance notation*	
			First	*Subsequent*
Optimal	$n-1$	'optimal'	x_1^*	$x_2^*, \ldots, x_{n-1}^*$
Heuristic	2	'optimal'	$\tilde{x}_1$	$\tilde{x}_2$
Heuristic	2	fitted	$\hat{x}_1$	$\hat{x}_2$

increase in cost from using the heuristic policy in lieu of the optimal policy, and end this section with a brief investigation into the sensitivity of the performance of the heuristic policy $\hat{\mathbf{x}}$ to distributional misspecification.

Define $V(\mathbf{x}|\boldsymbol{\zeta})$ to be the weighted waiting time, given (scaled) vectors of job allowance times $\mathbf{x}$ and realized service times $\boldsymbol{\zeta}$

$$V(\mathbf{x}|\boldsymbol{\zeta}) \doteq \sum_{i=1}^{n-1}(x_i - \zeta_i) + \sum_{i=2}^{n}\alpha_i w_i, \tag{20}$$

where

$$\begin{aligned} w_i &= (w_{i-1} + \zeta_{i-1} - x_{i-1})^+ \quad i = 2, \ldots, n,\\ w_1 &= 0. \end{aligned}$$

Then the expected relative increase in cost will be

$$\frac{E_\zeta[V(\hat{\mathbf{x}}|\boldsymbol{\zeta}) - V(\mathbf{x}^*|\boldsymbol{\zeta})]}{E_\zeta[V(\mathbf{x}^*|\boldsymbol{\zeta})]}, \tag{21}$$

which are estimated for each of the 210 test cases by taking the average over 10 000 000 generated realizations of $\boldsymbol{\zeta}$. The results are graphed in Fig. 7 for five representative values of n, for α between 0.01 and 1.0. The performance of the heuristic policy is uniformly excellent; its average cost is always within 2% of the cost of the optimal policy, and is generally within 0.5%.

Another performance measure of interest is the worst-case increase in relative cost: For which observation of service times $\boldsymbol{\zeta}$ would we most regret using $\hat{\mathbf{x}}$ instead of $\mathbf{x}^*$?

$$\frac{\max_{\zeta}[V(\hat{\mathbf{x}}|\boldsymbol{\zeta}) - V(\mathbf{x}^*|\boldsymbol{\zeta})]}{E_{\zeta}[V(\mathbf{x}^*|\boldsymbol{\zeta})]}. \tag{22}$$

It is straightforward to show that the numerator of Equation (21) will be maximized by $\boldsymbol{\zeta} = \mathbf{x}^*$, for which $V(\mathbf{x}^*|\mathbf{x}^*) = 0$; note that the denominator of (21) remains $E_{\zeta}[V(\mathbf{x}^*|\boldsymbol{\zeta})]$. These worst-case results are graphed in Fig. 8,

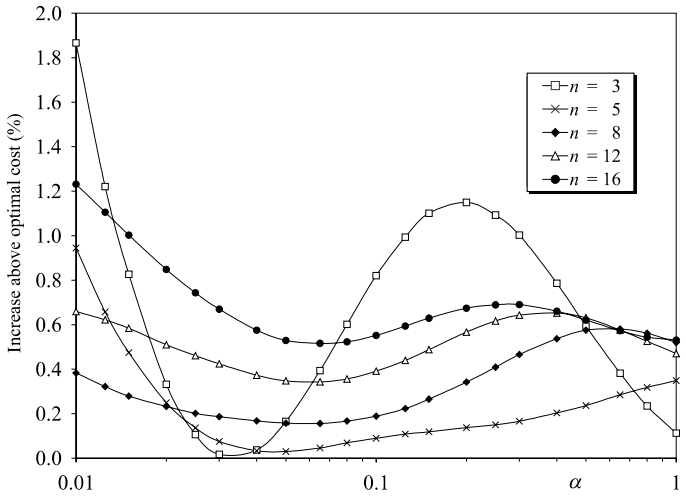

Fig. 7. Average heuristic performance: percent increase above optimal cost.

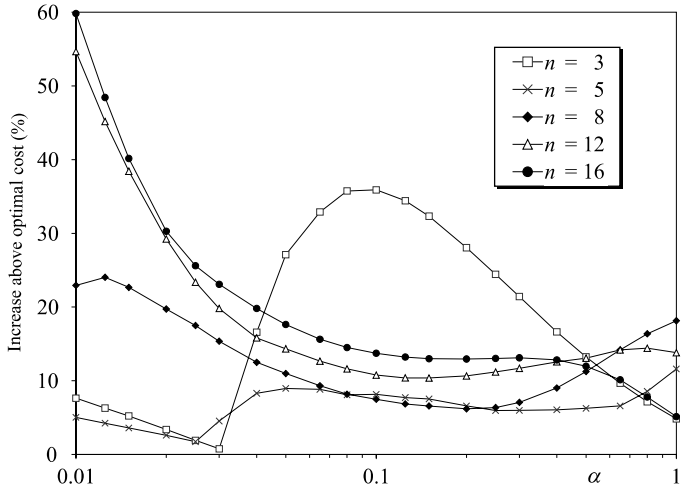

Fig. 8. Worst-case heuristic performance: percent increase above optimal cost.

again for α between 0.01 and 1.0 and for five representative values of n. The worst-case performance of the heuristic policy seems acceptable; it is never more than 60% above the average optimal cost, and for $n \geq 4$ and $\alpha \geq 0.04$ it is within 20% of the optimal cost.

So in terms of both the average and the worst-case performance, this simple closed-form atheoretic heuristic policy performs exceptionally well over the fairly broad test bed of parameters ($n \leq 16$, $0.01 \leq \alpha \leq 1.0$). This strong performance is somewhat surprising (albeit gratifying), given that both the form of the policy, and the values used within that policy, are approximated.

Up to now, all of the numerical results are based on the patient service times following the GLD fitted to the data set of Goldman *et al.* (1970) as shown in Fig. 1. Although this distribution is intuitively appealing (being unimodal and positively skewed), it is of course not the only choice. A thorough investigation into the effects of distributional misspecification for this problem would warrant its own paper, but it is important to develop some rough idea as to the robustness of this heuristic with regards to the distributional form of the service times. In order to do this, we examined its performance under three different service time distributions: the GLD's fitted to the data sets of Welch (1964) and of Brahimi and Worthington (1991), and the normal distribution. As with the data set of Goldman *et al.* (1970), the two GLD's were constructed through maximum likelihood estimation, and were scaled to yield a mean of zero and a standard deviation of one. (The alternative of sampling directly from the original histogram was not chosen, as it would have resulted in only 18 distinct observations for either data set.) The original histograms, and the fitted GLD's, are given in Figs. 9 and 10. Summary statistics (skewness and kurtosis) for all four distributions are given in Table 2, while the GLD parameter values are listed in Table 3. For comparative purposes, the distributions are graphed together in Fig. 11. It is clear that the data set of Brahimi and Worthington (1991) is considerably different than that of Goldman *et al.* (1970), with twice the skewness (1.292 versus 0.646) and decidedly more kurtosis (4.513 versus 3.678).

For testing purposes, we chose a single intermediate number of patients ($n = 8$), and found the optimal policies $\mathbf{x}^*$ for each of the three distributions for each of 21 different values of α. Again, we used $K = 10\,000$ observations and 1000 replications in calculating the optimal policy. We then used Equation (20) to calculate the average increase in costs from using the heuristic policy $\hat{\mathbf{x}}$ based on the data of Goldman *et al.* (1970), based on 10 000 000 generated observations of service times. The results are shown in Fig. 12; note that the solid line representing Goldman is the same as in Fig. 7, and is included as a baseline to indicate the average cost increase of using a heuristic policy, without distributional misspecification. Figure 12 shows that performance of the

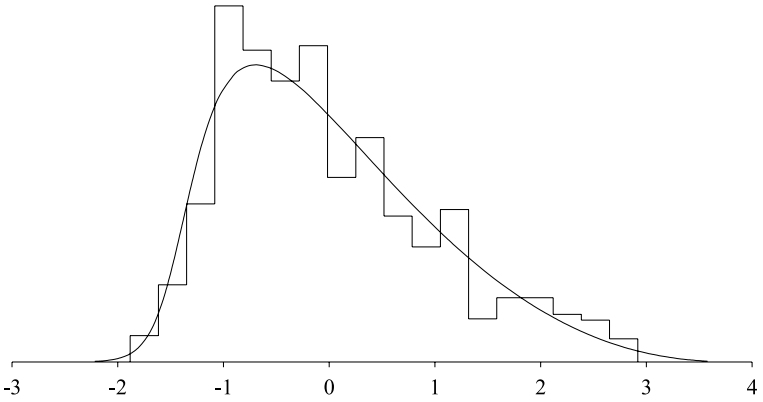

Fig. 9. Standardized distribution of patient service times, after Welch (1964).

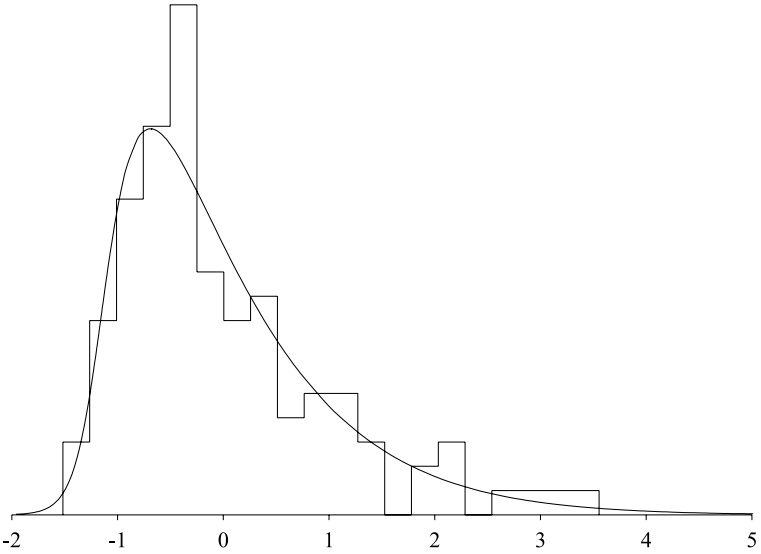

Fig. 10. Standardized distribution of patient service times, after Brahimi and Worthington (1991).

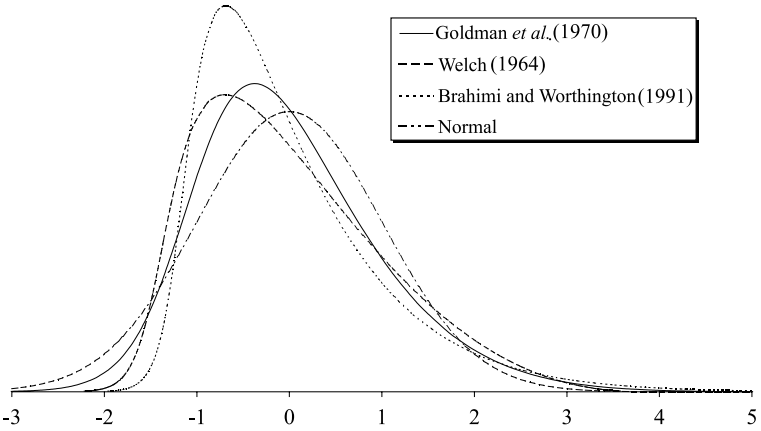

Fig. 11. Comparison of different standardized service time distributions.

heuristic policy given by Equations (16) and (17) is only slightly worse for different service time distributions, and generally remains under 1.5%. An exception is for small values of α, under the very skewed data of Brahimi and Worthington (1991), where the expected cost penalty approaches 7% above optimal. This is because small

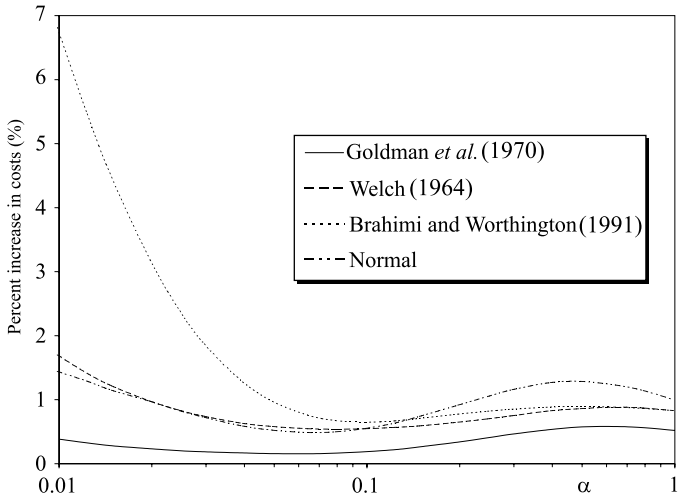

Fig. 12. Effects of distributional misspecification.

values of α will yield values of x_1^* near the very bottom of the left tail of the service time distribution. Because the left tail of Brahimi and Worthington is so much shorter than those of the other three distributions, values of x_1 derived from these other distributions (including the heuristic $\hat{x}_1$) will be fit poorly.

6. Conclusions and future research

The problem of setting appointment times is very important in many professional services, ranging from doctors and dentists to auto repair, particularly as a wider recognition is given to the importance of service quality as a means of competitive advantage. This model can also be applied to broader problems of scheduling physical resources over time, including the scheduling of an operating room and the setting of departure times for a train or an aircraft, for example. Even in the simplified case where the variability in service times dominates all other sources of uncertainty, the problem is quite intractable for more than two patients, requiring Monte Carlo integration or decomposition techniques.

In this paper we develop a very simple closed-form heuristic for setting job allowances (and appointment times) for up to 16 homoskedastic and equally-important (in terms of the value of their waiting time) customers. We show that this heuristic performs very well: usually averaging within 0.5% of the cost of the optimal policy, with worst-case performance usually within 20% of optimal. Although the service time distribution used for this heuristic was based on a single empirical study of surgery times, some limited testing shows that this heuristic performs quite well for other service times distributions, provided that their skewness is not so different as to drastically shift their lower tail.

There are a wide range of directions for possible future research. It is easy to extend this model to account for doctor tardiness (rather than idleness), or to include postponable tasks by the doctor (e.g., calling pharmacies, arguing with HMO's, etc.) which can be used to fill his or her idle time. Another generalization would allow for multiple (and interchangeable) doctors. Staying within the general model of this paper, other obvious extensions would be to expand the test bed to evaluate $\alpha > 1$, $\alpha < 0.01$, or $n > 16$, or to consider heteroskedastic customers, or customers with different waiting costs α_i.

As this research stream develops, it will be important to capture other uncertainties within the system. One such example would be the uncertainties of emergency and unscheduled appointments within the day. A first cut at this inclusion might replace the distribution of the length of a patient's service time with the multi-modal distribution of the length of the busy period of serving that customer and any and all emergency patients who preempt that patient.

Another important and significant source of uncertainty is the patient arrival times. The simplest model would be to model patient arrival times by a given distribution about his or her appointment time. A more realistic extension would also incorporate the difference between lateness and tardiness from the patient's perspective, where waiting may be perceived as less onerous prior to the scheduled appointment time. Modeling stochastic arrivals would allow us to evaluate the commonly-used 'block schedules', where multiple patients are assigned the same appointment time, and are seen by the doctor in the order of their arrival.

Microeconomic methodologies are being more frequently applied to problems of operations management. For this problem, an important application would be to use game theory to model patient arrival times. For example, if patients believe that their doctor is more likely to run behind schedule at the end of the day, they will not make much of an effort to arrive promptly at their appointment times. A formal model of this common behavior is apt to be quite rich.

References

Bailey, N.T.J. (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, Series B*, **14**, 185–199.

Bailey, N.T.J. (1954) Queuing for medical care. *Applied Statistics*, **3**, 137–145.

Blanco White, M.J. and Pike, M.C. (1964) Appointment systems in out-patients' clinics and the effect of patients' unpunctuality. *Medical Care*, **2**, 133–145.

Brahimi, M. and Worthington, D.J. (1991) Queuing models for out-patient appointment systems—a case study. *Journal of the Operational Research Society*, **42**, 733–746.

Charnetski, J.R. (1984) Scheduling operating room surgical procedures with early and late completion penalty costs. *Journal of Operations Management*, **5**, 91–102.

Denton, B. and Gupta, D. (2001) A sequential bounding approach for optimal appointment scheduling. Unpublished working paper, University of Minnesota.

Fishman, G.S. (1996) *Monte Carlo*: *Concepts*, *Algorithms*, *and Applications*, Springer-Verlag, New York, NY.

Fries, B.E. and Marathe, V.P. (1981) Determination of optimal variable-sized multiple-block appointment systems. *Operations Research*, **29**, 324–345.

Goldman, J., Knappenberger, H.A. and Shearson, W.T. (1970) A study of the variability of surgical estimates. *Hospital Management*, **110**, 46–46D.

Hall, R.W. (1991) *Queueing Methods for Services and Manufacturing*, Prentice-Hall, Upper Saddle River, NJ.

Halton, J. (1970) A retrospective and prospective survey of the Monte Carlo method. *SIAM Review*, **12**, 1–63.

Hammersley, J.M. and Handscomb, D.C. (1964) *Monte Carlo Methods*. Methuen, London.

Ho, C.-J. and Lau, H.-S. (1992) Minimizing total cost in scheduling outpatient appointments. *Management Science*, **38**, 1750–1764.

Jansson, B. (1966) Choosing a good appointment system—a study of queues of the type (*D*,*M*,1). *Operations Research*, **14**, 292–312.

Karian, Z.A. and Dudewicz, E.J. (2000) *Fitting Statistical Distributions*: *The Generalized Lambda Distribution and Generalized Bootstrap Methods*, CRC Press, Boca Raton, FL.

Liao, C.-J., Pegden, C.D. and Rosenshine, M. (1993) Planning timely arrivals to a stochastic production or service system. *IIE Transactions*, **25**, 63–73.

Mercer, A. (1960) A queuing problem in which the arrival times of the customers are scheduled. *Journal of the Royal Statistical Society, Series B*, **22**, 108–113.

O'Keefe, R. (1985) Investigating outpatient departments: implementable policies and qualitative approaches. *Journal of the Operational Research Society*, **36**, 705–712.

Pegden, C.D. and Rosenshine, M. (1990) Scheduling arrivals to queues. *Computers and Operations Research*, **17**, 343–348.

Ramberg, J.S. and Schmeiser, B.W. (1972) An approximate method for generating symmetric random variables. *Communications of the ACM*, **15**, 987–990.

Ramberg, J.S. and Schmeiser, B.W. (1974) An approximate method for generating asymmetric random variables. *Communications of the ACM*, **17**, 78–82.

Robinson, L.W., Gerchak, Y. and Gupta, D. (1996) Appointment times which minimize waiting and facility idleness. Unpublished working paper, DeGroote School of Business, McMaster University, Hamilton, Ontario, Canada.

Sabria, F. and Daganzo, C.F. (1989) Approximate expressions for queuing systems with scheduled arrivals and established service order. *Transportation Science*, **23**, 159–165.

Schafer, W.B. (1986) Keep patients waiting? Not in my office. *Medical Economics* **63**(10), 137–141.

Schmeiser, B. and Deutsch, S. (1977) A versatile four parameter family of probability distributions suitable for simulation. *AIIE Transactions*, **9**, 176–182.

Soriano, A. (1966) Comparison of two scheduling systems. *Operations Research*, **14**, 388–397.

Vanden Bosch, P.M. and Dietz, D.C. (2001) Scheduling and sequencing arrivals to an appointment system. *Journal of Service Research*, **4**, 15–25.

Wang, P.P. (1993) Static and dynamic scheduling of customer arrivals to a single-server system. *Naval Research Logistics*, **40**, 345–360.

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

Welch, J. (1964) Appointment systems in hospital outpatient departments. *Operational Research Quarterly*, **15**, 224–237.

Welch, J. and Bailey, N.T.J. (1952) Appointment systems in hospital outpatient departments. *The Lancet*, **1**, 1105–1108.

Weiss, E.N. (1990) Models for determining estimated start times and case orderings in hospital operating rooms. *IIE Transactions*, **22**, 143–150.

Biographies

Lawrence W. Robinson is an Associate Professor at the Johnson Graduate School of Management at Cornell University. He received his M.B.A. and Ph.D. degrees from the Graduate School of Business at the University of Chicago in 1986. His research focuses on problems of operating in an uncertain service environment; in particular, on developing practical heuristic policies that perform well and can be easily calculated. He is a member of INFORMS, ASQ, and Rotary.

Rachel R. Chen is currently a Ph.D. student at the Johnson Graduate School of Management at Cornell University. Her research focuses on supply chain management and the scheduling of service operations. She has been a member of INFORMS since 1999.

Contributed by the Feature Applications and Technology Management Department