MANAGEMENT SCIENCE
Vol. 54, No. 3, March 2008, pp. 565–572
ISSN 0025-1909 | EISSN 1526-5501 | 08 | 5403 | 0565

DOI 10.1287/mnsc.1070.0802

Scheduling Arrivals to Queues: A Single-Server Model with No-Shows

Refael Hassin, Sharon Mendel
School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel
{hassin@post.tau.ac.il, sharonmendel@hotmail.com}

Queueing systems with scheduled arrivals, i.e., appointment systems, are typical for frontal service systems, e.g., health clinics. An aspect of customer behavior that influences the overall efficiency of such systems is the phenomenon of no-shows. The consequences of no-shows cannot be underestimated; e.g., British surveys reveal that in the United Kingdom alone more than 12 million general practitioner (GP) appointments are missed every year, costing the British health service an estimated £250 million annually. In this study we answer the following key questions: How should the schedule be computed when there are no-shows? Is it sufficiently accurate to use a schedule designed for the same expected number of customers without no-shows? How important is it to invest in efforts that reduce no-shows—i.e., given that we apply a schedule that takes no-shows into consideration, is the existence of no-shows still costly to the server and customers?

Key words: appointment scheduling; health care; no-show; queue; single server
History: Accepted by Michael Fu, stochastic models and simulation; received October 6, 2006. This paper was with the authors $4\frac{1}{2}$ months for 4 revisions. Published online in *Articles in Advance* January 25, 2008.

1. Introduction

Queueing systems with scheduled arrivals are known as appointment systems. A better-designed appointment system can reduce waiting time for customers and increase the utilization of expensive personnel and other resources. An important aspect of customer behavior that influences the overall efficiency of such systems is the phenomenon of no-shows. Appointment systems of all kinds experience "disappointments" incidences on a regular basis. Empirical studies of no-shows cited by Cayirli and Veral (2003) indicate that in many health clinics the volume of no-shows strongly affects the system's performance. Some medical websites (Medical News Today 2004) quote figures of up to about 40% of no-shows. A British survey published in 2003 (BBC News 2003) reveals that in the United Kingdom alone more than 12 million general practitioner (GP) (family doctor) appointments are missed every year, costing the British health services an estimated £250 million annually. A survey performed by KPMG Consulting supports these figures (New Health Network 2002). A survey published in 2006 (Developing Patient Partnerships 2006, Guardian Unlimited 2006) reveals that these figures have not decreased over the years, and over 11 million GP appointments, and just over 5 million practice nurse appointments, are missed every year just in the United Kingdom. Sources in the American Air Force are also found to complain about no-shows (see American Air Force 2006). The above is just a drop in the ocean of practical evidence that no-shows are a crucial factor in the efficient performance of appointment systems, and specifically in the performance of national health services (NHS) all over the world. Nevertheless, despite the extensive work done on appointment systems, the issue of no-shows has been studied very little.

The first to present an extensive work dealing with appointment scheduling of outpatient services was Bailey (1952). A literature review on appointment policies can be found in the work of Mondschein and Weintraub (2003). Cayirli and Veral (2003) provide a comprehensive survey of research on appointment scheduling in outpatient services. They classify works by the research methodology used: analytical, simulation-based, and case study. A more general bibliography of queues in health care was provided by Preater (2001).

As a base point for our study, we use an analytical model for scheduling arrivals to queues where all customers show up (Pegden and Rosenshine 1990). We also refer to Stein and Cote (1994), who based their study on Pegden and Rosenshine (1990). They use the base model to study larger systems and also to study and compare the results obtained when adding a constraint of equally spaced scheduled appointments.

The first to address appointment systems with no-shows in an analytic approach was Mercer, who studies the nonequilibrium distribution of the queue length and also gives the results for the equilibrium distribution (see Mercer 1960, 1973). More recent are the works of Hutzschenreuter (2005) and of Kaandorp and Koole (2006). The former compares the performance of a selection of appointment-scheduling rules under D + noise/M/1 models, using simulation for solving instances of them. Kaandorp and Koole (2006) define a local search scheduling algorithm and prove that it converges to the optimal schedule with respect to a weighted sum of the average expected waiting times of customers, idle time of server, and tardiness in schedule. The algorithm is flexible and can incorporate no-shows.

This paper analyzes the scheduling of finite appointment systems with no-shows by providing a mathematical formulation of such a model with a single server and analyzing the effect of no-shows on the performance of the system. We aim to answer the following key questions:

- How should the (optimal) schedule be computed when there are no-shows?
- How is the efficiency of the system affected if the scheduled arrivals are restricted to be equally spaced?
- How important is it to reduce no-shows, i.e., given that we apply a solution that takes no-shows into consideration, is the existence of no-shows still costly to the server and customers?

Our results indicate that no-shows greatly affect the schedule and should be taken into account when designing an appointment system.

We study two types of appointment systems: In §2, time intervals between scheduled arrivals are not necessarily equal. A customer can be scheduled to arrive at any time after, or even at the same time, as the customer scheduled to arrive before her. In §3, we study systems with equal time intervals between scheduled arrivals. We study each type of system separately, and also compare the results obtained for each type. Further results and details can be found in Mendel (2006).

2. General Model

We assume that each customer has a probability of p of showing up, and if he shows up he does so exactly at his scheduled arrival time. The objective is to determine a schedule for a fixed number of customers, minimizing the sum of expected customers' waiting costs and expected server's availability cost. There is a single server to provide a Markovian service who remains available until the last customer leaves the system. The server has no prior knowledge of which customers will show up; hence, he has to remain available at least until the time the last customer is scheduled to arrive. Customers are served in the order of their scheduled appointments. Denote the scheduled arrival of the kth customer by t_k, then if $t_i = t_j$, $i < j$, and both customers i and j show up, customer i is served before customer j.

We assume that service times follow an exponential distribution and are identically distributed. For appointment systems in health services this identical distribution assumption is often reasonable because the extent of the ailment is not known at the time the appointment is made. A possible extension for further research may be multiple class models, where appointments are drawn from different distributions. Such an extension should involve a different solution method, because there is a sequencing problem involved.

2.1. Notation

n: Number of customers to be scheduled.

c_w: Customer's waiting cost per unit of time.

c_s: Server's availability cost per unit of time.

γ: Relative cost measure $\gamma = c_s/(c_s + c_w)$.

$1/\mu$: Mean service time (service exponentially distributed).

x_i: Time interval between the scheduled arrival times of the ith and the $(i+1)$st customers.

$\bar{x}$: A vector of intervals between scheduled arrivals $\bar{x} = (x_1, x_2, \ldots, x_{n-1})$.

t_i: Time of the ith scheduled arrival. $t_1 = 0$ and $t_i = \sum_{j=1}^{i-1} x_j$, $i = 2, \ldots, n$.

w_i^s: Expected waiting time in the queue of the ith customer if he shows up. $w_1^s = 0$.

p: Probability of a customer to show up.

$N(t_i)$: Number of customers in the system just prior to the time of the ith scheduled arrival.

$S(n, p)/M/1$: A queueing system with n scheduled independent customers, each showing up with probability p according to a specified schedule $S(n, p)$, and a single Markovian server.

2.2. Objective Function

The objective is to determine a vector $\bar{x}$ minimizing the sum of the expected customer waiting costs and server availability cost:

$$\Phi^1(\bar{x}) = c_w p \sum_{i=1}^{n} w_i^s + c_s\left(\sum_{i=1}^{n-1} x_i + E[\text{server's time after } t_n]\right)$$
$$= c_w p \sum_{i=1}^{n} w_i^s + c_s\left[\sum_{i=1}^{n-1} x_i + w_n^s + \frac{p}{\mu}\right]. \quad (1)$$

In this expression we use the fact that the expected server's time after t_n is the sum of the time he has to serve customers who showed up before t_n and are still in the system at t_n, which is the amount of time the last customer would have waited if he had showed up, and the expected service time of the last customer if he shows up. The objective function can be simplified by omitting the constant $c_s \cdot p/\mu$. We also divide by $(c_s + c_w p)$ to obtain that the objective to be minimized is

$$\Phi(\bar{x}) = (1 - \tilde{\gamma}) \sum_{i=2}^{n-1} w_i^s + \tilde{\gamma} \sum_{i=1}^{n-1} x_i + w_n^s, \quad (2)$$

where $\tilde{\gamma} = c_s/(c_s + c_w p)$ is the *relative server's availability cost*, or obtained by using the relative cost measure $\tilde{\gamma} = \gamma/(\gamma + p \cdot (1-\gamma))$.

To evaluate the effect of no-shows only on the variable costs, i.e., isolating the expected service time from the objective because it is "inevitable" and moreover a constant that does not depend on the schedule, we define $\Omega^*(\bar{x})$ as the expected total cost of waiting in the system. It consists of the sum of the expected cost of customers waiting in the queue and the idle time of the server. $\Omega^*(\bar{x})$ can be obtained from the objective function (1) by $\Omega^*(\bar{x}) = \Phi^1(\bar{x}) - E[\text{total service cost}] = \Phi^1(\bar{x}) - c_s pn/\mu$. Rewriting and simplifying the equation, we obtain

$$\Omega^*(\bar{x}) = \Phi(\bar{x}) - \frac{\tilde{\gamma} p(n-1)}{\mu}. \tag{3}$$

2.3. Recursive Expression for w_i^s

In this section we develop a general expression for w_i^s as a function of $\bar{x}$. The expected waiting time in the queue for a customer who shows up depends upon the number of customers he encounters upon arrival: $w_i^s \equiv \sum_{j=1}^{i-1} j/\mu \cdot \Pr\{N(t_i) = j\}$. The probability that there are j customers in the system at t_i depends upon whether or not the $(i-1)$th customer shows up. For $1 \le j < i$ and $i \ge 2$,

$$\begin{aligned}\Pr\{N(t_i) = j\} = p \cdot \sum_{k=0}^{i-j-1} & \Pr\{N(t_{i-1}) = j+k-1\} \\ & \cdot \Pr\{k \text{ departures between } t_{i-1} \text{ and } t_i\} \\ & + (1-p) \cdot \sum_{k=0}^{i-j-2} \Pr\{N(t_{i-1}) = j+k\} \\ & \cdot \Pr\{k \text{ departures between } t_{i-1} \text{ and } t_i\}.\end{aligned}$$

Because the service is Markovian, the probability of k departures between the $(i-1)$st and the ith scheduled arrivals (assuming there are at least k customers in the system at t_{i-1}) is the probability of exactly k events in a Poisson process with the rate of μ. Thus,

$$\begin{aligned}&\Pr\{N(t_i) = j\} \\ &\quad = p \cdot \sum_{k=0}^{i-j-1} \Pr\{N(t_{i-1}) = j+k-1\} \cdot \frac{(\mu x_{i-1})^k}{k!} e^{-\mu x_{i-1}} \\ &\qquad + (1-p) \cdot \sum_{k=0}^{i-j-2} \Pr\{N(t_{i-1}) = j+k\} \cdot \frac{(\mu x_{i-1})^k}{k!} e^{-\mu x_{i-1}} \\ &\quad = \sum_{k=1}^{i-j-1} \Pr\{N(t_{i-1}) = j+k-1\} \cdot \frac{(\mu x_{i-1})^{k-1}}{(k-1)!} e^{-\mu x_{i-1}} \\ &\qquad \cdot \left(\frac{p\mu x_{i-1}}{k} + 1 - p\right) + p \cdot \Pr\{N(t_{i-1}) = j-1\} \cdot e^{-\mu x_{i-1}} \\ &\qquad\qquad 1 \le j < i, \ \ i \ge 2.\end{aligned}$$

Similarly, the probability that the system is empty just prior to t_i for $i \ge 2$ is

$$\begin{aligned}&\Pr\{N(t_i) = 0\} \\ &\quad = p \cdot \sum_{k=1}^{i-1} \Pr\{N(t_{i-1}) = k-1\} \\ &\qquad \cdot \Pr\{\text{time between } t_{i-1} \text{ and } t_i \text{ suffices for at least } k \text{ departures}\} \\ &\qquad + (1-p) \cdot \sum_{k=0}^{i-2} \Pr\{N(t_{i-1}) = k\} \\ &\qquad \cdot \Pr\{\text{time between } t_{i-1} \text{ and } t_i \text{ suffices for at least } k \text{ departures}\} \\ &\quad = p \cdot \sum_{k=1}^{i-1} \left(\Pr\{N(t_{i-1}) = k-1\} \cdot \sum_{l=k}^{\infty} \frac{(\mu x_{i-1})^l}{l!} e^{-\mu x_{i-1}}\right) \\ &\qquad + (1-p) \cdot \sum_{k=0}^{i-2} \left(\Pr\{N(t_{i-1}) = k\} \cdot \sum_{l=k}^{\infty} \frac{(\mu x_{i-1})^l}{l!} e^{-\mu x_{i-1}}\right) \\ &\quad = \sum_{k=0}^{i-2} \Pr\{N(t_{i-1}) = k\} \\ &\qquad \cdot \left(1 - \sum_{l=0}^{k-1} \frac{(\mu x_{i-1})^l}{l!} e^{-\mu x_{i-1}} - p \cdot \frac{(\mu x_{i-1})^k}{k!} e^{-\mu x_{i-1}}\right).\end{aligned}$$

2.4. Solution Method

A closed-form solution for this optimization model exists only for $n = 2$ customers, a case of limited practical interest; for details and results see Mendel (2006). For larger systems we must obtain a solution numerically. The objective function used by Pegden and Rosenshine (1990) is believed to be convex despite the appearance of nonconvex terms, although no one has been able to prove the convexity. Each member in the sum that forms the no-shows model objective function (2) is a linear combination of a member of Pegden and Rosenshine's objective function; hence, (2) is convex if and only if Pegden and Rosenshine's objective function is convex. Based on this assumption, we obtain what is believed to be an optimal solution by applying sequential quadratic programming (SQP); see Mendel (2006) for details.

2.5. Solution Analysis

2.5.1. Schedule. As p decreases, appointments are scheduled closer together. From a certain value, some customers at the beginning of the schedule are scheduled to arrive at the same time. The interval width increases for the first few customers, then stays almost constant until it decreases for the last few customers. With no-shows, this phenomenon expands, and as p decreases and the relative server's availability cost increases, not only are the first few customers scheduled to arrive together, but so are the last few customers. The latter phenomenon occurs for relatively

Figure 1 **Interarrival Times** $S(10, 0.70)/M/1$

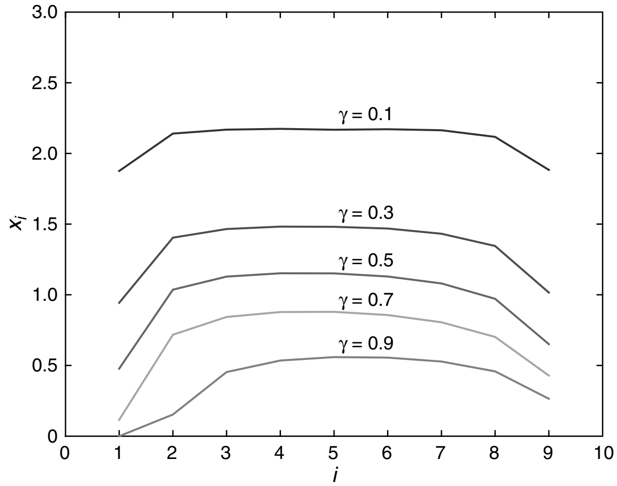
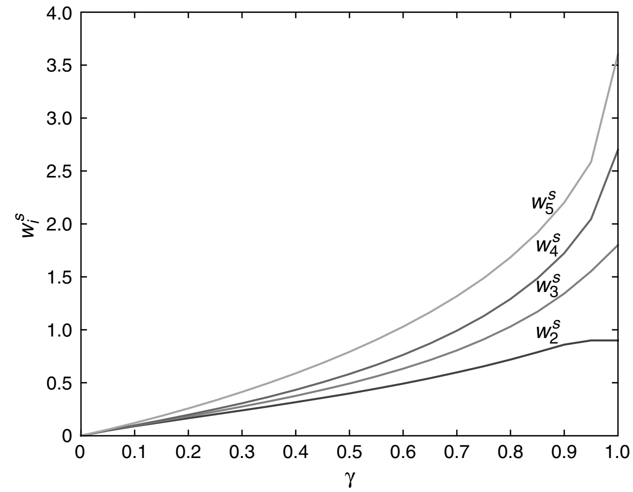


Figure 3 **Expected Waiting Times** $S(5, 0.90)/M/1$



low values of p. For example, in a system with 10 customers, we begin to observe this pattern only for $p \leq 0.30$ and $\gamma \geq 0.90$. Figures 1 and 2 present the spacing between scheduled arrivals for various values of γ and 10 customers. The interval numbers are given on the horizontal axis and the spacings between scheduled arrivals are given on the vertical axis. Point (i, x_i) on a certain γ graph in these figures is the value found for x_i in the relevant model, i.e., the scheduled interarrival time between customer i and customer $i+1$ for a system of $S(10, p)/M/1$, with p and γ as stated in the figure.

The intuitive explanation for these spacings is similar to that given by Stein and Cote (1994). The last few customers are scheduled to arrive closer, or even together, to avoid the server being idle while only a few customers remain to arrive. The scheduling of the first few customers to arrive close or even together fits what is known as Bailey's Law (Bailey 1952), which recommends scheduling the first customers together to later reduce the server's idle time.

2.5.2. Customers' Expected Waiting Times. Our objective function is a combination of expected customers' waiting times and server's availability time. However, if the expected waiting times are too high, the system's management may consider reevaluating the relative cost, so greater consideration would be given to waiting cost. Figures 3 and 4 present expected waiting times of customers who show up, for various relative server's availability costs. The variance of the expected waiting times of customers who show up is quite high, and the expected waiting time of a customer who shows up increases as his scheduled place in line increases, i.e., $w_{i+1}^s > w_i^s \ \forall\, i = 1, \ldots, n-1$. This could be explained by the queue's discipline. As more customers are scheduled to arrive together, the more they wait if more than one of them arrives. We also note that the

Figure 2 **Interarrival Times** $S(10, 0.30)/M/1$

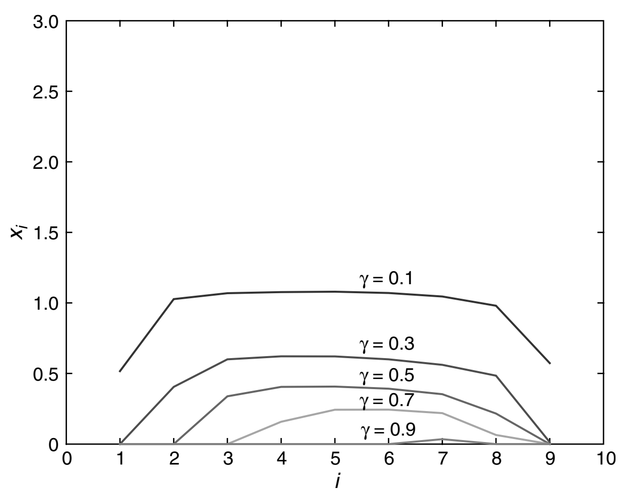
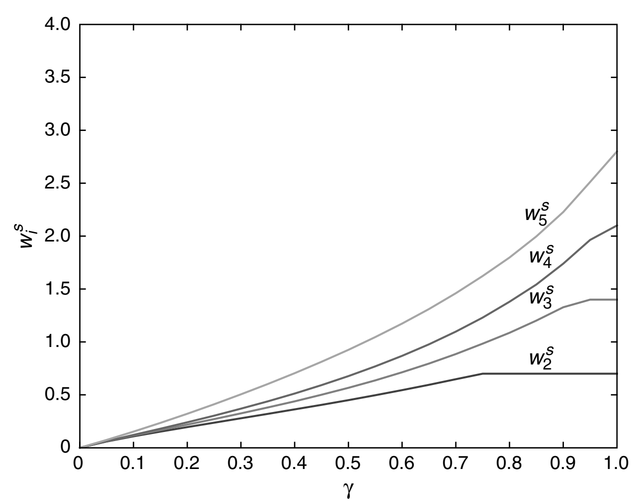


Figure 4 **Expected Waiting Times** $S(5, 0.70)/M/1$



expected waiting time for customers who show up increases with p, i.e., if fewer customers are expected to arrive, the customers who show up are less likely to wait due to service of earlier customers. Nevertheless, these customers are waiting much longer than they would if they were served by a system that serves the same expected number of customers and is designed and operates as an appointment system where all customers show up.

2.5.3. Quantifying the Impact of No-Shows on Customers' Expected Waiting Times. Comparing the expected waiting times for customers who show up for a $S(n, p < 1.0)/M/1$ model to the $S(n', 1.0)/M/1$ model where $np = n'$, we find that if all customers who do not show up would have notified in advance that they were not going to show up, a schedule could have been designed with smaller expected waiting times for those who show up.

Figures 5 and 6 each present a comparison between the expected waiting times for customers who show up in a $S(n, p < 1.00)/M/1$ model and the same expected waiting times of the equivalent $S(n', 1.00)/M/1$ model where $np = n'$. In these graphs the relative server's availability cost, γ, is given on the horizontal axis, and the expected waiting times measures of customers who show up are given on the vertical axis. The waiting times of the models with no-shows are drawn in solid lines, whereas the ones of the models where all customers show up are drawn in dashed lines. We note that as p decreases, the impact of the no-shows increases. If a large portion of customers do not show up, the expected waiting time of the customers who do show up is much higher than the expected waiting time they would have had if the schedule was originally designed only for them. The influence of the no-shows phenomenon is greater

Figure 5 **Waiting Measures Comparison** $S(8, 0.625)/M/1$ **and** $S(5, 1.0)/M/1$

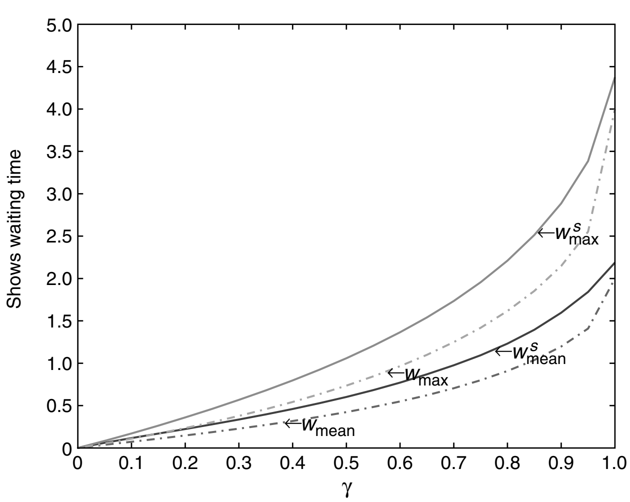
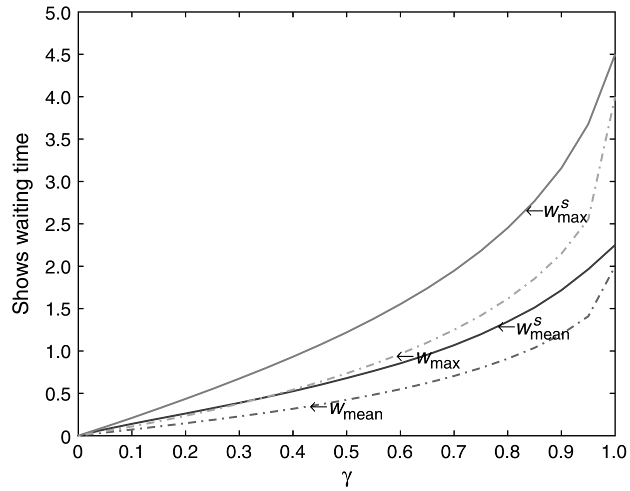


Figure 6 **Waiting Measures Comparison** $S(10, 0.50)/M/1$ **and** $S(5, 1.0)/M/1$



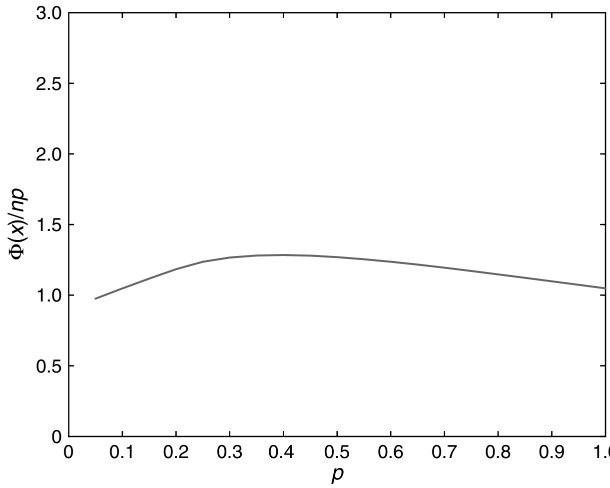
for smaller relative server's availability cost, meaning that the impact of no-shows is more severe when the importance given to customers' waiting times, to minimize them, is higher.

2.5.4. Objective Function's Value. For high γ, the minimum value found for the objective function for $p = 1$, where all customers show up, is higher than for some other showing up probabilities. Assuming the relative cost γ is given and a system's operating costs are all represented in (2), the operating costs can be brought to a lower expected value by manipulating a specific p portion of the customers to show up.

2.5.5. Operational Costs per Customer. Two other measures of the system that are of much interest concern operational costs per customer. These measures are defined with respect to $\Phi(\bar{x})$ and $\Omega(\bar{x})$ by dividing them by np. $\Omega(\bar{x})/np$ is a measure for system's waiting costs per customer, whereas $\Phi(\bar{x})/np$ also takes into account the service costs. Figures 7 and 8 present the values of these measures as a function of p for various values of γ. For both measures, the maximum value they obtain is not necessarily when all customers show up. When comparing the behavior of the total costs with those of the costs per customer, we find that their picks are not obtained at the same showing-up probability, and moreover, the picks of the operational measures graphs are obtained for a lower p. This may imply that customers who show up "pay" more if many of the scheduled customers do not show up.

2.5.6. Cost of No-Shows. To evaluate the cost of no-shows, we compare the costs of systems where all customers show up to systems with no-shows with the same expected number of customers who show up.

Figure 7 $\Phi(\bar{x})$ **per Showing Customer** $S(10, p)/M/1$, $\gamma = 0.70$



Figures 9 and 10 each present a comparison between the costs of a $S(n, p < 1.00)/M/1$ model and the costs of the equivalent $S(n', 1.00)/M/1$ model where $np = n'$. In these graphs the relative server's availability cost is given on the horizontal axis, noted by γ, and the costs are given on the vertical axis. The costs of the models with no-shows are drawn in solid lines, whereas the costs of the models where all customers show up are drawn in dashed lines. We note that as the probability of customers not showing up increases, the cost of no-shows increases. The influence of the phenomenon on the costs is greater on a smaller relative server's availability cost. These findings conform with the findings detailed in §2.5.3 for the expected waiting times. Nevertheless, the impact of no-shows on the expected waiting times, measured in percentages, is not as high. Thus, the no-shows also increase the server's idle time.

Figure 8 $\Omega(\bar{x})$ **per Showing Customer** $S(10, p)/M/1$, $\gamma = 0.30$

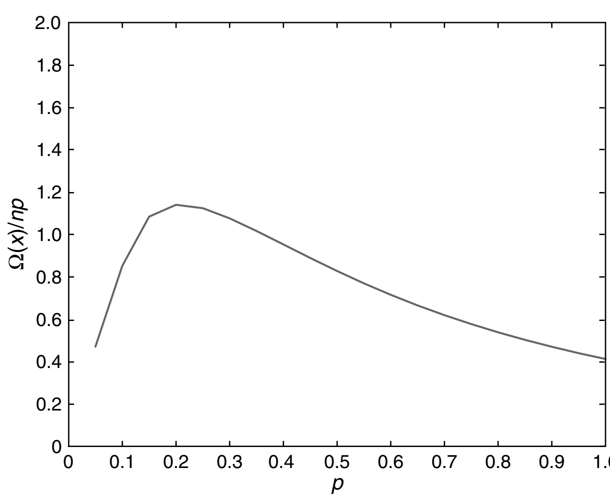
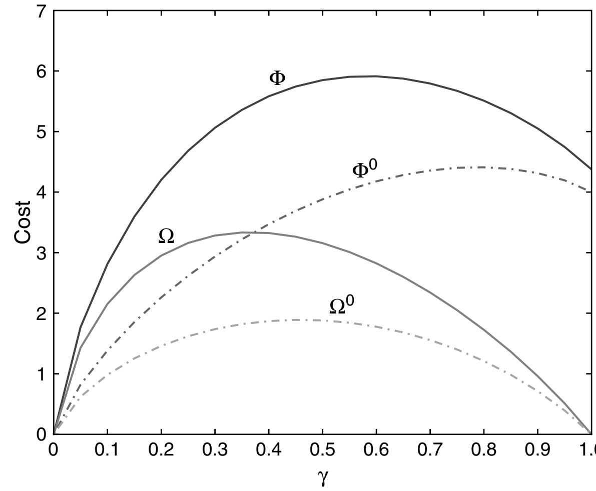


Figure 9 **Costs Comparison** $S(8, 0.625)/M/1$ **and** $S(5, 1.0)/M/1$



3. The Equally Spaced Model

Following the work of Stein and Cote (1994), we also consider the case where the customers are scheduled to arrive at the system at equally spaced times, i.e., $x_1 = x_2 = \cdots = x_{n-1}$. We add to their model the attribute that each customer shows up with a probability $p \in (0, 1]$.

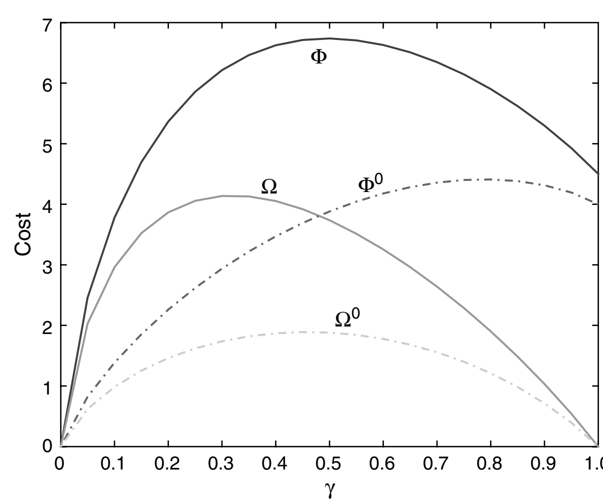
As stated by Stein and Cote, the equally spaced model is of interest because it provides a realistic restriction to the scheduling problem, because in most appointment systems the appointments are scheduled using a fixed interval between scheduled arrivals.

3.1. Solution Method

The equations representing our model remain the same, with the exception that $x_i = x_{eq}$ $\forall\, i$. Thus, based on (2) our objective is to minimize

$$\Phi_{eq}(x_{eq}) = (1 - \tilde{\gamma}) \sum_{i=2}^{n-1} w_i^s + \tilde{\gamma}(n-1)x_{eq} + w_n^s. \qquad (4)$$

Figure 10 **Costs Comparison** $S(10, 0.50)/M/1$ **and** $S(5, 1.0)/M/1$



The solution methods are similar to those used in the unrestricted model. Even though we are now looking for the value of a single variable x_{eq} that optimizes the objective function (4) (as opposed to a vector of values), there is still no closed-form solution. Hence, as in the unrestricted model, we obtain the solution numerically for the simplified case of $\mu = 1$ by using SQP methods.

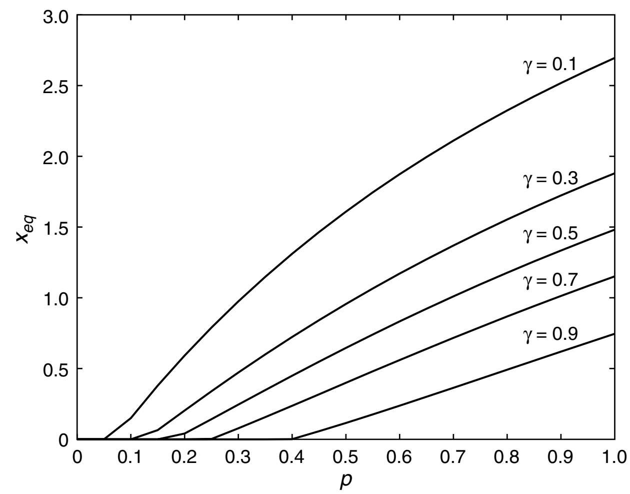
3.2. Analysis of the Equally Spaced Solution

As found by Stein and Cote for the basic model, we also find that for a system with no-shows, the effect of adding a constraint to force equally spaced intervals does not materially change the value of the objective function for any combination of p and γ, and determines an interval that is approximately the average of the equivalent unrestricted model. Also, the expected waiting times in the equally spaced model behave in a similar manner to the expected waiting times in the unrestricted model, and have similar values.

3.2.1. Relation Between the Equal Spacing Solution and the Showing-Up Probability. From a practical point of view, it is of interest to study the relation between the equal spacing solution and the showing-up probability. Assuming a given server's availability relative cost, the equal spacing solution depends upon the showing-up probability. We find that for a wide range of relative cost values there seems to be an almost linear relation between the equal spacing solution and the showing-up probability. This can be noted in Figures 11 and 12, which present, for a given number of customers, the equal spacing solution for given relative costs as a function of the showing-up probabilities.

3.2.2. No-Shows' Impact on the Equal Spacing Solution. A comparison between the equal spacing solution of systems where all customers show up

Figure 12 **Equal Spacing Solutions** $S(10, p)/M/1$



to systems with no-shows with the same expected number of showing-up customers, reveals that the spacing with no-shows is smaller, and it decreases as the showing probability p decreases. Figures 13 and 14 compare the equal spacing solution of $S(n, p < 1.00)/M/1$ models and the corresponding $S(n', 1.00)/M/1$ models where $np = n'$. The relative server's availability cost, γ, is given on the horizontal axis, and the spacing is given on the vertical axis. The spacings with no-shows are drawn in solid lines, whereas those of the models where all customers show up are in dashed lines. If customers would have notified in advance that they were not going to show up, a more spacious schedule could have been designed for those who do show up. This aligns with the impact of no-shows on the expected waiting times and costs, as detailed in §§2.5.3 and 2.5.6. Due to expected no-shows, the schedule is more condensed, with longer expected waiting times leading to

Figure 11 **Equal Spacing Solutions** $S(5, p)/M/1$

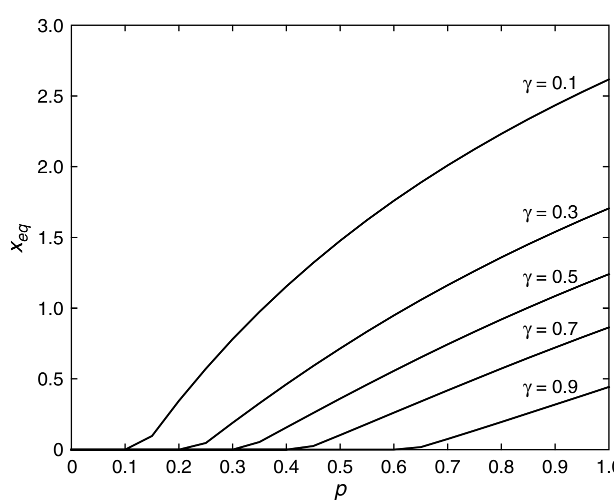


Figure 13 x_{eq} $np = 8$ **Comparisons**

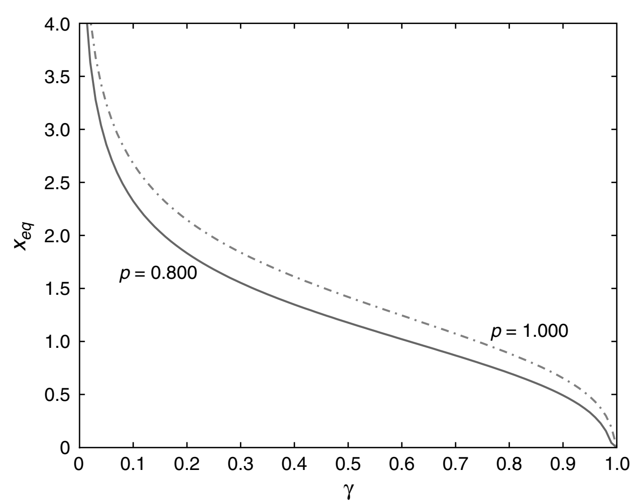
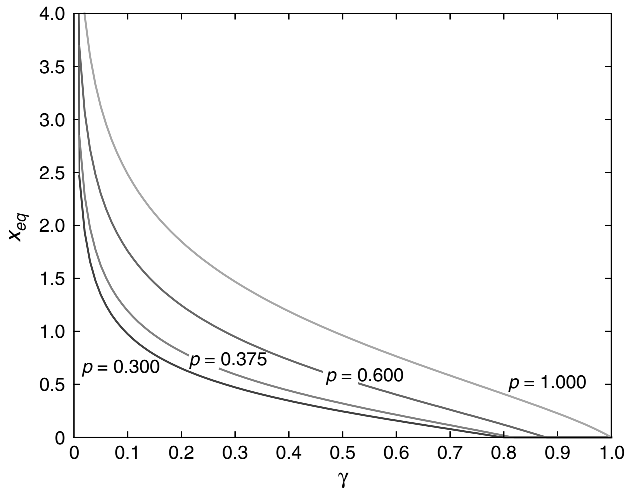


Figure 14 x_{eq} $np = 3$ **Comparisons**



higher costs than there would have been if the system was designed only for the customers who do show up eventually. We also note that the impact of no-shows on the spacing is less for extreme relative service availability costs, i.e., the impact when γ is very small or very large is not as notable as it is for intermediate values of γ. In these extreme cases the schedule is highly influenced by the relative cost; hence, the impact of no-shows is not as significant.

Acknowledgments
This research was supported by the Israel Science Foundation (Grant 237/02).

References

American Air Force. 2006. No-shows for appointments wastes time, money. Retrieved July 1, 2007, http://public.travis.amc.af.mil/news/story.asp?storyID=123017095.

Bailey, N. 1952. A study of queues and appointment systems in hospitals outpatients departments with special reference to waiting times. *J. Roy. Statist. Soc.* **14** 185–199.

BBC News. 2003. GPs back "no show" fines. Retrieved July 1, 2007, http://news.bbc.co.uk/1/hi/health/3187101.stm.

Cayirli, T., E. Veral. 2003. Outpatient scheduling in health care: A review of literature. *Production Oper. Management* **12** 519–549.

Developing Patient Partnerships. 2006. Missed appointments increasing despite same day appointments. Retrieved July 1, 2007, http://www.dpp.org.uk/en/1/pr2006kioci.qxml.

Guardian Unlimited. 2006. Unwelcome break. Retrieved July 1, 2007, http://society.guardian.co.uk/health/story/0,,1856663,00.html.

Hutzschenreuter, A. 2005. Queueing models for outpatient appointment scheduling. Thesis, Faculty for Mathematics and Economics, Ulm University, Germany.

Kaandorp, G., G. Koole. 2006. Optimal outpatient appointment scheduling. Working paper, Department of Mathematics, Vrije University, Amsterdam.

Medical News Today. 2004. Why 42% of clinic appointments are no-shows: An issue of respect. Retrieved July 1, 2007, http://www.medicalnewstoday.com/medicalnews.php?newsid=17238.

Mendel, S. 2006. Scheduling arrivals to queues: A model with no-shows. M.Sc. thesis, Department of Statistics and Operations Research, Tel Aviv University. Retrieved July 1, 2007, http://www.math.tau.ac.il/~hassin/sharon_thesis.pdf.

Mercer, A. 1960. A queueing problem in which the arrival times of the customers are scheduled. *J. Roy. Statist. Soc., Ser. B* **22** 108–113.

Mercer, A. 1973. Queues with scheduled arrivals: A correction, simplification and extension. *J. Roy. Statist. Soc., Ser. B* **35** 104–116.

Mondschein, S. V., G. Y. Weintraub. 2003. Appointment policies in service operations: A critical analysis of the economic framework. *Production Oper. Management* **12** 266–286.

New Health Network. 2002. What patients really think of the NHS. Retrieved July 1, 2007, http://www.newhealthnetwork.co.uk/template3.asp?ID=41.

Pegden, C. D., M. Rosenshine. 1990. Scheduling arrivals to queues. *Comput. Oper. Res.* **17** 343–348.

Preater, J. 2001. A bibliography of queues in health and medicine. Keele Mathematics Research Report 01-1, Keele University, Staffordshire, UK.

Stein, W. E., M. J. Cote. 1994. Scheduling arrivals to queues. *Comput. Oper. Res.* **21** 607–614.