

Clinic Overbooking to Improve Patient Access and Increase Provider Productivity*

Linda R. LaGanga[†]

Director of Quality Systems, Mental Health Center of Denver, 4141 East Dickenson Place, Denver, CO 80222, e-mail: linda.laganga@colorado.edu

Stephen R. Lawrence

Leeds School of Business, University of Colorado at Boulder, 419 UCB, Boulder, CO 80309-0419, e-mail: stephen.lawrence@colorado.edu

ABSTRACT

The problem of patient *no-shows* (patients who do not arrive for scheduled appointments) is significant in many health care settings, where no-show rates can vary widely. No-shows reduce provider productivity and clinic efficiency, increase health care costs, and limit the ability of a clinic to serve its client population by reducing its effective capacity. In this article, we examine the problem of no-shows and propose appointment overbooking as one means of reducing the negative impact of no-shows. We find that patient access and provider productivity are significantly improved with overbooking, but that overbooking causes increases in both patient wait times and provider overtime. We develop a new clinic utility function to capture the trade-offs between these benefits and costs, and we show that the relative values that a clinic assigns to serving additional patients, minimizing patient waiting times, and minimizing clinic overtime will determine whether overbooking is warranted. From the results of a series of simulation experiments, we determine that overbooking provides greater utility when clinics serve larger numbers of patients, no-show rates are higher, and service variability is lower. Even with highly variable service times, many clinics will achieve positive net results with overbooking. Our analysis provides valuable guidance to clinic administrators about the use of appointment overbooking to improve patient access, provider productivity, and overall clinic performance.

Subject Areas: Appointment Scheduling, Health Care Policy, No-shows, Overbooking, Service Operations, Scheduling Policies and Systems, Simulation, and Utility Assessment.

INTRODUCTION

Efficient use of clinic resources is critical in an era of rapidly escalating health care costs and calls for improved health care efficiency (Sweeney, 1996), so health care managers face increasing pressure to control costs while delivering high-quality

*We gratefully thank the editor, associate editors, and two anonymous reviewers for their valuable comments and suggestions during the review process.

[†]Corresponding author.

care (Wright, Bretthauer, & Côté, 2006). The problem of patient *no-shows* (patients who do not arrive for scheduled appointments) is significant in many health care settings, where no-show rates can vary from as little as 3% to as much as 80% (Rust, Gallups, Clark, Jones, & Wilcox, 1995). No-shows reduce provider productivity and clinic efficiency, increase health care costs, and limit the ability of a clinic to serve its client population by reducing its effective capacity.

In this article, we examine the problem of no-shows and propose appointment overbooking as a means of reducing the negative impact of no-shows. We find that patient access and provider productivity are significantly improved with overbooking, but that overbooking causes increases in both patient wait times and provider overtime. The relative values that a clinic assigns to (i) serving additional patients, (ii) minimizing patient waiting times, and (iii) minimizing clinic overtime will determine whether overbooking is warranted; these trade-offs are captured by a new clinic utility function developed herein. Our analysis provides valuable guidance to clinical administrators for expanding their capacity to serve patients while managing the real and implicit costs of clinic operations and patient wait times.

Our research is motivated by a publicly funded outpatient community mental health center seeking to maximize its capacity to treat underserved members in the community while managing limitations imposed by sharply reduced funding from state and federal sources. In this clinic, almost 30% of adult patients failed to show up for their scheduled appointments with psychiatrists. Consistent with findings reported in prior literature (Chesanow, 1996; Larkin, 1999; Murray & Berwick, 2003), the clinic found that no-shows reduce provider utilization and productivity while negatively affecting customer satisfaction and quality of health care. In one related study, only 74% of surveyed community mental health services consumers were satisfied with their access to services (Colorado Department of Human Services, 2005). In addition to community mental health centers, the no-show problem may be particularly severe for pediatric clinics, hospitals, and neighborhood medical and dental clinics (Bean & Talaga, 1995).

While prior appointment scheduling literature has tangentially studied overbooking in conjunction with patient no-shows (Blanco White & Pike, 1964; Vissers, 1979; Rohleder & Klassen, 2002), previous research has not explicitly studied the impact and promise of patient overbooking on clinic performance. To our knowledge, the research reported in this article is the first to focus on the use of overbooking as a means to mitigate the negative impact of no-shows, improve patient access, and increase provider productivity in clinical settings.

We contribute to the health care management literature in four ways. First, we show how scheduling complexity increases when appointment overbooking is used to compensate for no-shows. By demonstrating the dynamics of patient arrival uncertainty in both the timing and the number of no-shows, we differentiate clinic overbooking from overbooking for revenue management in transportation services. Second, we introduce a new analytic utility model that evaluates appointment overbooking in terms of trade-offs between the benefits of serving additional patients and the costs of increased patient wait time and provider overtime. This utility model enables an administrator to tailor our results to the specific characteristics of a clinic. Third, with simulation experiments, regression analysis, and sensitivity

experiments, we show that appointment overbooking in health care clinics can have a significantly positive net impact on clinic performance by increasing patient access and improving clinic productivity, which translates into reduced clinic costs and improved patient satisfaction and outcomes. Fourth, we provide managerial insights into the practical use of appointment overbooking in real-world clinics and demonstrate its application in a large publicly funded mental health clinic. We also identify situations in which overbooking is most likely to be beneficial and, conversely, in which it is likely to be counterproductive.

Although this article is motivated by problems in health care clinics, our research is relevant to any setting in which service provider productivity and customer access to services are impacted by problematic customer no-shows, including professional services, such as law offices and clinics; government offices; retail services, such as tax return preparation offices and salons; counseling centers; and admissions offices, among many others.

The remainder of the article is organized as follows. We first review previous literature relevant to our research and demonstrate that overbooking has not been extensively studied as a means of mitigating the negative impact of no-shows. In the third section, we develop a formal model of clinic utility that incorporates the costs of no-shows and the benefits of overbooking; we also develop analytic results for small clinics. In the fourth section, we report on the results of an extensive simulation study that examines the costs and benefits of no-shows and overbooking in a variety of clinic settings. These results generalize to a large range of clinic sizes and patient no-show rates. In the fifth section, we examine the sensitivity of our results to several of our modeling assumptions and provide insights into how our results can be implemented in practice. Finally, we offer concluding remarks and directions for future research.

OVERVIEW OF APPOINTMENT SCHEDULING

Scheduling service operations has been the subject of scholarly investigation for some years (Easton & Goodale, 2005), and within this broad domain, the scheduling of health care operations has received considerable attention (Wright, Bretthauer, & Côté, 2006). Appointment scheduling has been examined in the research literature for some decades (Bailey, 1952; Ho & Lau, 1992; Klassen & Rohleder, 1996). Other authors have contributed to the literature on service-operation no-shows from a variety of disciplines and perspectives, including medical practice, health care administration, operations management, marketing, and transportation planning. Little work has been reported, however, on the use of overbooking to mitigate the negative impact of no-shows in appointment-oriented services such as clinical health care.

In contrast, transportation revenue management has been extensively examined (Barnhart, Belobaba, & Odoni, 2003; Van Ryzin & Talluri, 2003; Lieberman, 2004, 2005), where overbooking has been studied in terms of capacity utilization and profitability using perishable asset revenue management (Weatherford & Bodily, 1992; Toh & Raven, 2003). However, appointment overbooking is very different from transportation services overbooking, because appointment no-shows are spread over time, while transportation no-shows all occur at a single point

in time. This difference in problem structure requires quite different solution approaches to the problem of no-shows.

In the domain of appointment scheduling, field studies have found large variations in appointment no-show rates among medical specialties and geographic regions (Sharp & Hamilton, 2001) and among patient populations (Garuda, Javalgi, & Talluri, 1998). A number of studies have reported patient no-show rates as low as 10% (Brahimi & Worthington, 1991; Warden, 1995). Sharp and Hamilton (2001) reported a 12% no-show rate at outpatient clinics in the United Kingdom. According to Barron (1980), eight studies at inner-city clinics, community health centers, and university medical centers indicate no-show rates of 10–30% while the estimated no-show rates for private practice are 2–15%. An even wider range of no-show rates, 3–80%, is reported in a study by Rust et al. (1995) of 200 public pediatric clinics.

Health care researchers and some practitioners have focused on finding the causes of no-shows and eliminating or reducing them. They have considered costs such as analysis of patients and their behavior, as well as the implementation costs of programs or practices to boost patient attendance rates (Shonick & Klein, 1977; Bean & Talaga, 1995; Garuda et al., 1998; Campbell, Chez, Queen, Barcelo, & Patron, 2000).

Reported reasons for no-shows include lack of transportation, scheduling problems, oversleeping or forgetfulness, and lack of child care (Campbell et al., 2000). The probability of patient no-shows might relate to factors such as patient age, gender, and number of previous appointments (Shonick & Klein, 1977); appointment lead time (Grunebaum, Luber, Callahan, Leon, Olfson, & Portera, 1996); and/or Medicaid status (Rust et al., 1995). McCarthy, McGee, and O'Boyle (2000) and Sharp and Hamilton (2001) suggest that no-show rates might increase if expected patient wait times grow too long at a clinic for arriving patients. Approaches that have been successfully applied to reduce no-shows include sending patients reminder cards (Rust et al., 1995), calling patients to remind them of appointments, and providing information about public transportation (Bean & Talaga, 1995).

Operations management and statistical methods have been employed in some studies of clinical appointment scheduling systems that measure performance as the weighted sum of patient wait time and provider idle-time costs (Bailey, 1952; Welch & Bailey, 1952; Ho & Lau, 1992). These articles identify no-shows as a significant factor in schedule performance and measure some of their effects, but do not focus on how to handle no-shows or reduce their negative impact in the scheduling system.

Out of 36 articles categorized in a recent review of outpatient scheduling literature by Cayirli and Veral (2003), only 11 include the possibility of no-shows and only 4 include policies to mitigate the effects of no-show behavior. Blanco White and Pike (1964) focus on the punctuality of patient arrivals as it impacts clinic performance, but only consider patient no-shows on a limited basis. Fetter and Thompson (1966) investigate the impact of walk-ins as a counterbalance to no-shows, but do not consider appointment overbooking. Vissers and Wijngaard (1979) adjust the mean and variance of service times to compensate for both no-shows and walk-ins, but do not directly study overbooking. Vissers (1979) recommends

that the interval between scheduled appointments be reduced to compensate for no-shows, but provides no analysis or data to support this recommendation.

In other literature most closely related to this article, Shonick and Klein (1977) show how to use patient no-show probabilities, conditioned on patient characteristics related to no-show behavior, to overbook enough patients so that the expected number of arrivals is equal to the target number to be seen. They acknowledge the consequent problems of patient wait time and provider idle time, but do not use these implicit costs to evaluate performance and do not consider the possibility that overbooking can result in clinic overtime. More recently, Rohleder and Klassen (2002) examine rolling-horizon scheduling policies when demand for appointments fluctuates. They investigate the use of double-booking (a form of overbooking) and overtime to increase capacity when appointment demand loads are high, but the purpose of such double-booking is not to compensate for no-shows, which they hold constant at 5%. They evaluate alternative scheduling policies using performance measures of appointment lead time, patient wait time, clinic session length, provider utilization, and provider idle time.

Building on this previous research, we explicitly focus on the problem of no-shows and on the consequences of overbooking with the objective of helping clinical decision makers determine whether or not appointment overbooking should be used. Our appointment overbooking model accommodates a wide range of clinic sizes and no-show rates, allowing its application in a variety of clinical practices.

We also extend an assumption of previous research that demand for appointments is at least as great as supply, which implies that the number of scheduled appointments is at least as great as the clinic size. Overbooking in clinical appointment scheduling is motivated by the more specific assumption that demand for appointments exceeds supply. Therefore, overbooking provides another benefit in addition to increased provider utilization, which is the recovery of idle capacity from no-shows, thus allowing increased patient access and service. We further assume that pricing for health care services does not vary dynamically to promote the sale of excess capacity or to maximize revenues. These assumptions differentiate clinical appointment systems from other service operations that use yield management (Kimes, 1989) or transportation revenue management (McGill & Van Ryzin, 1999).

APPOINTMENT OVERBOOKING MODEL

We model a clinic that may have multiple providers (physicians, clinical staff, etc.), but in which each provider serves a dedicated set of patients (i.e., one provider does not serve the patients of another), so that each provider can be modeled in isolation. Each patient is seen for a fixed period of time D . A constant service time is consistent with many clinical-setting operations and allows us to focus on the uncertainty caused by no-shows rather than on any uncertainty introduced by service-time variation. For many other clinics, uncertain service times are the norm, so we demonstrate in our sensitivity analysis section that our results apply even when clinic service times are highly uncertain.

A clinical provider has the capacity to service N patients in a *clinical session*, where a *session* is defined as an uninterrupted block of clinic time such as a morning, afternoon, or perhaps an entire day. Parameter N also defines the clinic size, consistent with previous literature (Blanco White & Pike, 1964). The *duration* C of a clinical session is defined to be the duration of N consecutive sessions, or $C = ND$. At the conclusion of a clinical session, all service is terminated until the start of the next clinical session sometime later. The short duration of clinic sessions and the transient nature of clinic operations preclude the use of steady-state queuing models.

Patients are normally scheduled for appointments at intervals T throughout a clinic session. On average, a fraction R ($0 \leq R \leq 1$) do not arrive for their appointments and are no-shows, while the remaining fraction $S = 1 - R$ arrive (show) for their assigned appointments. An average of RN appointment slots are idle during a clinic session because of no-shows; this can substantially decrease provider productivity and increase the cost of clinic service.

To compensate for no-shows, the clinic has the option of overbooking a clinic session by scheduling more appointments K than it has capacity N (i.e., $K \geq N$). Following Vissers and Wijngaard (1979) and Vissers (1979), we assume that the time interval T between scheduled appointments is proportionally reduced by the number of sessions overbooked:

$$T = C/K = ND/K. \quad (1)$$

When a clinic overbooks a session, it runs the risk of having more patients x actually arrive during the session than its capacity N , so that $K \geq x \geq N$. In such cases, patient p may have to wait for a period W_p before being served by a provider. Overbooking may cause a provider to work to a time F beyond the nominal ending time C of a session (overtime), so that $F \geq C$.

Finally, we assume that the clinic incurs benefits and costs according to its appointment scheduling policy. We use nonfinancial utility measures (Metters & Vargas, 1999) for benefits and costs in our model, because providers in not-for-profit health care systems often value serving patients in need more than they value monetary benefits. Therefore, this study is relevant both for for-profit and not-for-profit health care providers. For each patient seen, the clinic receives a net marginal benefit π , which could be in the form of revenues received in a fee-for-service environment or as the increase of some utility measure in a not-for-profit setting, less the direct variable monetary, and/or utility costs of providing service. If patients are kept waiting for service, a marginal cost of ω is incurred per unit of wait time. This cost could be lost future revenues, decreases in goodwill, reduced patient utility, or other measures that value the patient's time. The marginal cost of a provider working overtime beyond the normal end time C of a clinic session is τ , which could include explicit labor overtime costs, costs of ancillary services and utilities, and the implicit costs of overworked staff and reduced employee morale, among others. Cost parameters π , ω , and τ are assumed to be denominated in the same units so that, for example, the ratio ω/τ represents the value of a patient's time relative to the provider's. Note that the numéraire of π , τ , and ω is immaterial to our analysis, but that their ratios (relative weights) are important.

Appointment Scheduling without Overbooking

To provide a baseline against which to compare the costs and benefits of overbooking, consider the expected utility of an appointment scheduling system in which there are no-shows but no overbooking. The *gross utility* U_G of a clinic session is $U_G = \pi N$, which represents the maximum benefit that accrues if all scheduled appointments show (are realized). We define *realized clinic utility* as $U_R = \pi SN = SU_G$, the expected benefit of a session after adjusting for expected no-shows. The utility cost of no-shows is the *reduction in utility* U_C , defined as:

$$U_C = U_G - U_R = \pi(1 - S)N = \pi RN. \quad (2)$$

The *marginal cost of no-shows* with respect to no-show rate R is simply πN . In the absence of overbooking, determining expected costs is straightforward because there is a direct linear relationship between show rate S and realized utility U_R , and between no-show rate R and the cost of no-shows U_C .

Appointment Scheduling with Overbooking

When patients are overbooked, scheduling complexity increases. A clinic benefits from greater provider utilization and increased utility revenues, but runs the risk of incurring patient waiting costs, overtime costs, or both. Patient wait-time costs are incurred when more patients arrive in a period of time than can be seen. Overtime costs are incurred when the arrival pattern of patients and/or the number of patient arrivals forces the clinic to work beyond its normal clinic duration in order to serve all arrived patients.

We define *overbooking utility* \bar{U}_N as the net utility gain (or loss) that a clinic expects to obtain from overbooking:

$$\bar{U}_N = \pi(\bar{A} - SN) - \omega\bar{W} - \tau\bar{O}, \quad (3)$$

where $\bar{A} = SK$ is the expected number of patients who arrive for scheduled appointments, $\bar{W} = \sum_p W_p/x$ is the expected waiting time for each of the x patients who show, and $\bar{O} = \bar{F} - C$ is expected clinic overtime (see Table 1). The first term represents the expected net benefit obtained by overbooking patients; the second term is the cost of changes in average patient waiting times, and the third term is the expected cost of increased overtime caused by overbooking.

Expected overbooking utility function \bar{U}_N (equation (3)) builds upon the performance measures of previous research that balance patient wait times with provider idle time (Bailey, 1952; Ho & Lau, 1992). For example, Ho and Lau (1992) use cost parameters to evaluate the trade-off between patient wait time and provider idle time that result from alternative scheduling rules. Utility function (3) differs from previous research in that we include the expected benefit of serving additional patients $\pi(\bar{A} - SN)$ due to overbooking, and the expected cost of clinic overtime $\tau\bar{O}$ that can arise from overbooking.

Note that the expression for overbooking utility \bar{U}_N implicitly includes the cost of server idle time. We assume that clinic personnel and facilities are available for the duration of a clinic session whether or not there are patient no-shows, and so their costs are fixed and sunk. Because they do not vary with nor influence the overbooking or scheduling policies employed, server costs are not explicitly modeled

Table 1: Notation.

$\bar{A} = KS$	Expected number of arriving patients
$C = ND$	Duration of a clinic
c_s	Coefficient of variation of provider service time (service variability)
D	Duration of an appointment
F	Finish time of a clinic after all arriving patients have been serviced
\bar{F}	Expected clinic finish time
K	Number of appointments scheduled for a clinic ($K \geq N$)
N	Number of sessions (appointments) in a clinic; clinic size; clinic capacity
\bar{O}	Expected clinic overtime
\bar{P}	Expected server productivity (utilization)
$R = 1 - S$	No-show rate of scheduled appointments ($0 \geq R \geq 1$)
$S = 1 - R$	Show rate of scheduled appointments ($0 \geq S \geq 1$)
$T = N/C$	Interval between scheduled appointments ($T \leq D$)
$U_C = U_G - U_R$	Utility cost of no-show patients
$U_G = \pi N$	Gross utility realized if all scheduled patients show (arrive)
U_N	Overbooking utility – net utility of arriving patients less penalties for patient wait times and clinic overtime when overbooking is employed
$U_R = SU_G$	Net utility of all patients who show (arrive)
W_p	Wait time of arriving patient p
\bar{W}	Expected wait time of all arriving patients
x	Number of patients who actually arrive for their appointments
π	Marginal net benefit of one additional patient
τ	Marginal cost or penalty of clinic overtime ($F > C$)
ϖ	Marginal cost or penalty of patient wait time (per patient)

in utility function (3). Nonetheless, server idle time and productivity are important managerial performance measures in health care. Many health care professionals consider provider productivity to be an important measure of clinic performance (Tonges, 1985; Chesanow, 1996; Baum, 2001; Chung, 2002; McCarthy, 2002; Cole, 2003). We define *provider productivity* P (or provider utilization) to be the ratio of the total time a provider is delivering service and the total length of the provider's work day including overtime, and use this definition in our subsequent analysis of the impact of overbooking on appointment scheduling performance in our simulation study section.

To demonstrate the dynamics of patient arrival uncertainty, several patient arrival patterns that illustrate both the benefits and the hazards of overbooking are shown in Table 2. In these examples, clinic capacity is five appointments ($N = 5$); patient show rate is 50% ($S = .5$), and each appointment lasts 1 time unit ($D = 1$). With overbooking, the clinic schedules 10 appointments for each session ($K = 5/.5 = 10$) and so compresses the time between scheduled appointments from 1 time unit to .5 time units ($T = .5$). With $S = .5$, the clinic expects only 5 of the 10 scheduled appointments to show up, on average.

Five cases were selected as examples of possible combinations of patient wait time and provider overtime in terms of each factor's occurrence or nonoccurrence. The Base Case shows the unique ideal arrival pattern in which neither patient wait time nor provider overtime occurs. Only 5 of 10 scheduled patients arrive, as expected, and the arrival pattern of patients is spaced such that no patient waits

Table 2: Impact of several patient show/no-show patterns.

	Regular Time										Overtime			
Time Slot	1	2	3	4	5	6	7	8	9	10	11	12	13	
Start Time	0	.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	
Base Case	<i>Expected number of patients (5) arrive, evenly spaced</i>										<i>No overtime</i>			
Arrivals	A_1	X_2	A_3	X_4	A_5	X_6	A_7	X_8	A_9	X_{10}				
Service	D_1		D_3		D_5		D_7		D_9					
Waiting	<i>No patients wait</i>													
Case 1	<i>Expected number of patients (5) arrive, bunched early</i>										<i>No overtime</i>			
Arrivals	A_1	A_2	A_3	X_4	A_5	X_6	A_7	X_8	X_9	X_{10}				
Service	D_1		D_2		D_3		D_5		D_7					
Waiting		W_2	W_3		W_5		W_7							
Case 2	<i>Expected number of patients (5) arrive, one late arrival</i>										<i>OT</i>			
Arrivals	A_1	X_2	A_3	X_4	A_5	X_6	A_7	X_8	X_9	P_{10}				
Service	D_1		D_3		D_5		D_7		I	D_{10}				
Waiting	<i>No patients wait</i>													
Case 3	<i>Expected number of patients (5) arrive, bunched late</i>										<i>OT</i>			
Arrivals	A_1	X_2	A_3	X_4	X_5	X_6	A_7	A_8	A_9	X_{10}				
Service	D_1		D_3		I	I	D_7		D_8					D_9
Waiting								W_8	W_9					
Case 4	<i>More patients arrive (6) than expected (5)</i>										<i>OT</i>			
Arrivals	A_1	A_2	A_3	X_4	A_5	X_6	A_7	X_8	A_9	X_{10}				
Service	D_1		D_2		D_3		D_5		D_7					D_9
Waiting		W_2	W_3		W_5		W_7		W_9					

A_i indicates an arriving patient who shows (arrives) for an appointment scheduled in time slot i .

X_i indicates a patient who is a no-show for an appointment scheduled in time slot i .

D_i indicates timeslots in which service occurs for patient i .

W_i indicates periods during which patient i is waiting for service.

I indicates periods during which the server (clinician) is idle.

No-show rate $S = .5$; clinic size $N = 5$; number booked $K = [N/S] = 10$.

Appointment duration $D = 1$; time interval between appointments $T = SD = .5$.

to be seen and no overtime is incurred. In Case 1, patient wait time occurs and provider overtime does not, illustrating the early-session arrival of a run of patients who subsequently must wait to be seen. However, fewer arrivals later in the session provide time for the clinic to work off its backlog and to avoid overtime. In Case 2, provider overtime occurs but patient wait time does not, because when each patient arrives, service has already been completed for all previous patients. The

Table 3: Expected net overbooking utility for small clinics.

N	K	Show Rate S	Net Overbooking Utility \bar{U}_N
1	1	$0.667 < S \leq 1$	$\bar{U}_N \leq 0$ because $K = N \Rightarrow \pi S(K - N) = 0$
2	2	$0.80 < S \leq 1$	
3	3	$0.833 \leq S \leq 1$	
1	2	$0.40 < S \leq 0.667$	$\bar{U}_N = \pi S - \omega S^2 D(1 - S)/2 - \tau [T(2 - S)(S + 1) - C]$
1	3	$0.167 \leq S < 0.5$	$\bar{U}_N = \pi S(3 - N) - \omega S^2[(D - T) + 0.5(1 - S)\max\{0, D - 2T\}] - \tau[3S^3 D + S^2(1 - S)(5D + \max\{2T, D\} + T) + 3S(1 - S)^2(T + D) - C]$
2	3	$0.5 \leq S < 0.833$	

arrival of a single patient late in a session causes overtime to occur, because the service duration extends past the clinic’s regular operating time. In Case 3, both patient wait time and provider overtime occur because of the late-session arrival of a run of patients who must wait for service and who force the clinic to work overtime. Case 4, like Case 3, illustrates the occurrence of both patient wait time and provider overtime and also shows that when more than the expected number of patients shows up, there is both patient wait time and overtime. There are, of course, many other possible arrival patterns for this simple clinic setting (there are $2^{10} = 1,024$ possible permutations), but these five cases serve to illustrate that with appointment overbooking, in contrast to transportation overbooking, the timing, as well as the number, of patient arrivals, can have a large impact on clinic performance measures.

Analytic Results for Small Clinics

To better understand the dynamics of overbooking, we analyze overbooking cost results for several clinics of size $N \in \{1, 2, 3\}$ patients. We limit our analysis to $K \in \{1, 2, 3\}$, because the number of patients booked K is equal to the ratio of clinic size N and show rate S rounded to the nearest integer, and K can assume any positive integer value as show rate S declines toward zero. Table 3 shows the expected overbooking utility \bar{U}_N for feasible combinations of N and K . Derivations of these results are included in the Appendix.

As examples, consider two small clinics of size $N = 2$. Each has the same cost structure with marginal patient benefit $\pi = 1$, marginal patient wait cost $\omega = 1.5$, and marginal overtime cost $\tau = 2$. The first clinic has a show rate of 60% and the second a show rate of 80%. By the results of Table 3, both clinics should consider overbooking with $K = 3$ patients scheduled for each clinic session. The interval between appointments for both clinics is $T = N/K = 2/3$ with $D = 1$. Then, from the overbooking utility function for $S = .6$ in Table 3,

$$\bar{U}_N = 1(.6) - 1.5(.12) - 2(.136) = .148, \quad \text{Clinic 1}$$

so $\bar{U}_N > 0$ and overbooking is recommended. But for the second utility with $S = .8$,

$$\bar{U}_N = 1(.8) - 1.5(.2133) - 2(.592) = -.704, \quad \text{Clinic 2}$$

so $\bar{U}_N < 0$ and overbooking is not recommended. This result indicates it is impossible to make general statements about overbooking in specific clinical settings

without knowing both the pattern of patient no-show behavior and the cost structure of the clinic.

SIMULATION STUDY OF OVERBOOKING

To verify the insights of the previous section and to explore larger problems in which the number of patients to be booked is larger than $K = 3$, we undertook a series of simulation experiments to investigate the impacts and opportunities of overbooking in larger clinic settings.

Experimental Design

To determine the expected overbooking utility and other performance parameters resulting from overbooking, we conducted a full factorial simulation study using clinic size N and show rate S as the principal experimental factors. Patient wait times W , server productivity P , and session finish times F were the dependent variables measured during the simulations. Consistent with Vissers (1979) and Ho and Lau (1992), we used five clinic size levels $N \in \{10, 20, 30, 40, 50\}$ spanning a range of clinics from those in which service providers see only about one patient every hour to those in which providers serve a patient every 10 minutes. Ten factor levels for show rates $S \in \{1.0, .9, \dots, .1\}$ were included in order to span the entire practical range of patient arrival percentages (Rust et al., 1995). A total of 10,000 replications were completed for each of the 50 factor combinations, for a total of 500,000 ($5 \times 10 \times 10,000$) observations. In pilot studies preceding the main experiment, the 95% confidence half-widths for both session finish times F and patient wait times W_p were no more than 2% of the point estimates.

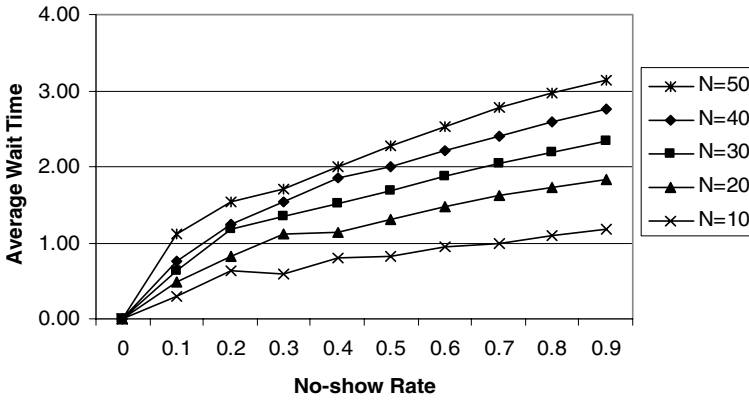
Within each observation, the number of scheduled appointments K was calculated as

$$K = [N/S], \quad (4)$$

because the target number of patients to be served is $N = SK$, and the ratio was rounded to the nearest integer. Although a clinic could overbook more or less than K appointments as calculated in equation (4), our focus in this article is the operational cost and value of attempting to recover (through overbooking) the service capacity loss caused by no-shows. Evaluating the full impact of no-shows in terms of the trade-off between the cost and the value of overbooking to compensate for them, on average over many clinic sessions, requires that the number of appointments be $K = N/S$. Service durations were arbitrarily set to $D = 1$ for ease of analysis. This assumption does not impact the implications of our simulation because its results are scaleable (Bailey, 1952). Thus, from the proportional reduction of the time interval T between appointments to fit the overbooked appointments into the clinic session as shown in equation (1) and illustrated in Table 2, $T = N/K$.

For each iteration of the simulation, the clinic was divided into K appointment slots, each with a duration of T time units. One patient was scheduled for each of the K appointment slots. Patients arrived for service with independent show probability S exactly at their scheduled arrival times. Arriving patients waited until the provider was free (if busy) and then were served with duration D . Waiting patients were seen in the order of their scheduled appointments. In cases where patients were no-shows and there was not a backlog of waiting patients, the service

Figure 1: Simulation results for mean patient wait time W with overbooking as a function of no-show rate R and clinic size N .



Note: Without overbooking, mean patient wait times are zero in all cases.

provider was idle. In cases where more patients arrived than could be served during the duration C of the clinic, the clinic worked overtime in order to serve all waiting patients. Because there was no variability in service time, overbooking was the sole cause of any patient wait time or provider overtime that occurred, thus we could attribute the costs in our utility function directly to the practice of overbooking. Data were collected for patient arrivals and waiting times, for provider idle time, and for clinic overtime. It was not necessary to specify costs or utility measures prior to conducting the simulation experiments—marginal utility benefits π , marginal patient wait costs ω , and marginal overtime costs τ were applied *ex post* to the simulation results.

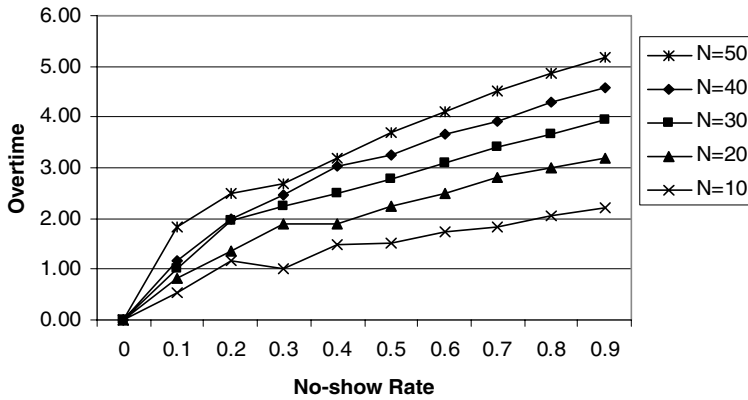
Experimental Results

Results of the simulation experiments show a number of interesting patterns that can be used to guide clinic scheduling and management. First, for all clinic sizes, overbooking causes both expected patient wait time \bar{W} and expected clinic overtime \bar{O} to increase as no-show rates R increase. Second, expected provider productivity \bar{P} improves with overbooking as no-show rates increase. And third, overbooking increases expected overbooking utility \bar{U}_N for most (but not all) clinics as clinic size increases. These results have important managerial implications for the effective management of health care clinics. Each is described in more detail in the following and is illustrated in Figures 1–5.

Wait time and overtime

Figures 1 and 2 respectively show average patient wait time \bar{W} and average clinic overtime \bar{O} as a function of no-show rate R and clinic size N . Note that, without overbooking, patients never wait and clinic overtime is never incurred, as shown in these figures for the case with no overbooking. These figures demonstrate that with overbooking, both patient wait time and clinic overtime increase as the

Figure 2: Simulation results for mean clinic overtime O with overbooking as a function of no-show rate R and clinic size N .



Note: Without overbooking, mean clinic overtime is zero in all cases.

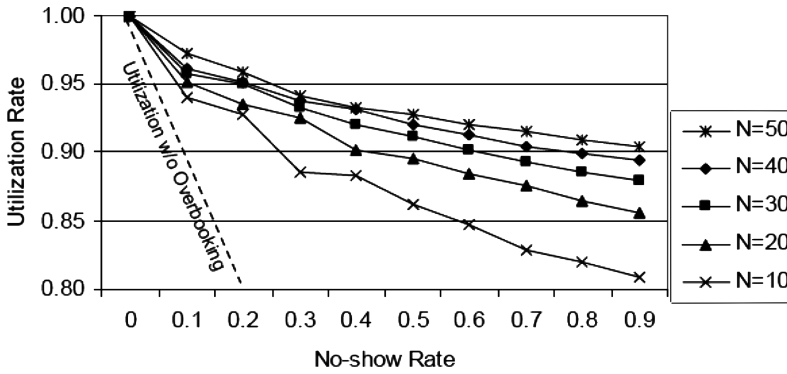
patient no-show rate increases. This occurs because of the effects of patient no-show patterns. As demonstrated in Table 2, both the frequency and pattern of patient no-shows can have a negative impact on patient wait times and clinic overtime. When too many overbooked patients arrive and/or patients arrive in problematic patterns, waiting and overtime costs are incurred. However, when too few patients arrive for a clinic session or when they arrive in benign patterns, no positive waiting or overtime benefits accrue to the clinic to balance the costs of problematic situations. The one-sided nature of this cost structure means that overbooking will always increase average patient waiting time and will always increase average clinic overtime, as Figures 1 and 2 suggest. The managerial implication of this result is that even in the presence of no-shows, clinic administrators must be willing to endure increased patient waiting times and increased clinic overtime if patient overbooking is employed. Further, waiting and overtime are disproportionately greater for small clinics than for large ones.

Provider productivity

Without overbooking, provider productivity is equal to the patient show rate S and so declines linearly with increasing no-show rates. Figure 3 shows that with overbooking, provider productivity declines as no-show rates increase, but at a rate that is much lower than without overbooking. With overbooking, average provider productivity never drops below 80% even when the no-show rate is 90%, whereas provider productivity (utilization) would only be 10% without overbooking. This result demonstrates that an overbooking policy can provide robust productivity performance results even in the presence of large no-show rates. Figure 3 also shows that with overbooking, provider productivity is greater for larger clinics than for smaller ones due to the portfolio effect of larger clinics.

The managerial implication of this result is that overbooking can dramatically improve clinic resource productivity. Efficient use of clinic resources is critical in

Figure 3: Simulation results for mean provider productivity P (utilization) with overbooking as a function of no-show rate R and clinic size N .



Note: The dashed line shows that expected provider utilization without overbooking is equal to show rate S .

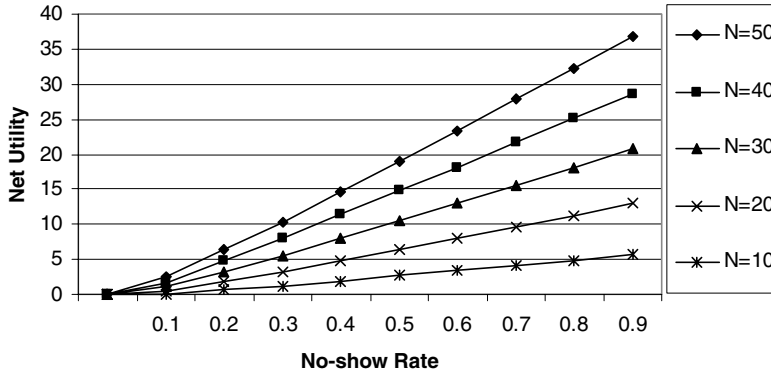
an era of rapidly increasing health care costs and resulting cost-control initiatives (Sweeney, 1996). Overbooking can clearly help to improve overall clinic utility as well as the productivity of individual providers. Improved productivity translates to increased availability of services, which in turn leads to improved patient outcomes. In a study of psychiatric patients after hospital discharge, patients were shown to be more likely to attend their follow-up outpatient appointments when they were scheduled within two weeks of discharge, and timely follow-up care reduced the likelihood of costly hospital readmission (Kruse & Rowland, 2002). Improved productivity of clinic providers helps reduce overall health care costs to the community.

Net overbooking utility

Net overbooking utility U_N measures the aggregate benefit (or cost) of serving more patients, offset by the cost of increased patient waiting times and increased clinic overtime, relative to running a clinic without overbooking (equation (3)). Note that without overbooking, net utility U_N is zero in all cases, because U_N only measures the benefits (losses) that derive from overbooking. Also, U_N implicitly captures the benefits of increased clinic productivity through the benefits of serving additional patients.

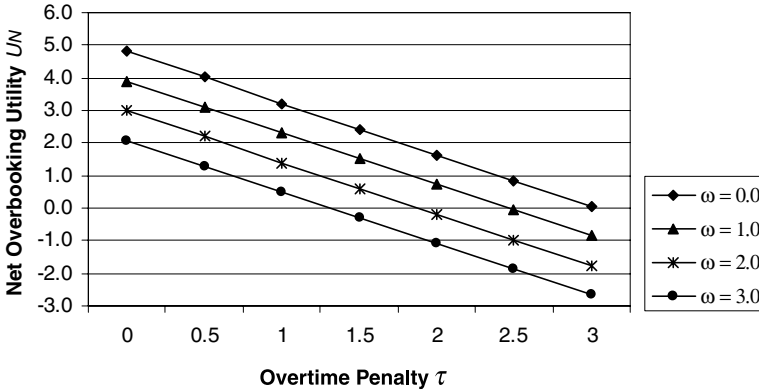
Figure 4 shows the overbooking utility patterns for a realistic clinic setting in which the marginal benefit of an additional patient π is larger than patient waiting costs ω , but slightly smaller than clinic overtime costs τ ; specifically, $\pi = 1$, $\omega = .5$, and $\tau = 1.2$. These parameter values were determined in consultation with the administrators of a large, public mental-health clinic. These administrators put a high value on serving as many patients as possible, but were concerned with the impact excessive overtime could have on clinicians and staff. The administrators were also interested in controlling patient waiting times, but gave this a lower priority than serving as many patients as possible. Simulation results for these

Figure 4: Simulation results for mean net overbooking utility U_N as a function of no-show rate R and clinic size N , with utility parameters $\pi = 1$, $\omega = .5$, and $\tau = 1.2$.



Note: Without overbooking, mean net utility is zero in all cases by definition.

Figure 5: Average net overbooking utility U_N as a function of patient waiting penalty ω and clinic overtime penalty τ with the benefit of an additional patient fixed at $\pi = 1.0$, clinic size $N = 16$, and no-show rate $R = .30$.



Note: From (5) and (6), expected waiting time $\bar{W} = .91D$ and expected overtime $\bar{O} = 1.58D$.

parameter values indicate overbooking is beneficial for all no-show rates R and all clinic sizes, as shown in Figure 4.

The utility of overbooking will, of course, depend on the relative weighting of parameters π , ω , and τ , as shown in Figure 5. From utility function (3), it is clear that in cases in which the patients' waiting time cost is much smaller than the benefit of seeing additional patients ($\omega \ll \pi$) and/or the implicit costs of clinic overtime are much smaller ($\tau \ll \pi$), then the net overbooking utility will be positive and overbooking will be beneficial. Alternatively, in cases in which the patients'

waiting time cost is much larger than the benefit of seeing additional patients ($\omega \gg \pi$) and/or the implicit costs of clinic overtime are much larger ($\tau \gg \pi$), then net overbooking utility will be negative and overbooking will not be warranted. Finally, in the majority of cases in which one parameter does not dominate, the expected benefits of overbooking must be calculated using equation (3). Figure 5 illustrates how net overbooking utility varies with parameters π , ω , and τ . Depending on the relative weighting of these parameters, net utility can be positive, marginal, or negative. The managerial implication of this result is that clinic administrators must carefully and thoughtfully develop appropriate parameter values for improved patient access π , patient waiting ω , and clinic overtime τ in order to evaluate the efficacy of overbooking for a particular clinic.

Clinic size effects

Figures 1–3 indicate that clinic size N has a significant impact on patient waiting times W , the propensity for overtime O , provider productivity P , and, consequently, on overbooking utility U_N , as shown in Figure 4, which indicates that the impact of no-show rate on overbooking utility increases with clinic size. Generalizing, smaller clinics benefit less from overbooking than do larger clinics. This is because the portfolio effects of larger clinics reduce the variance in patient waiting times and clinic overtime because of the greater number of patients seen during a clinic session. In small clinics, one arriving (or no-show) patient can have a significant impact on average patient waiting times and/or on clinic overtime. In larger clinics, a difference of one patient has a proportionally smaller effect. Additionally, the larger number of patients arriving at large clinics provides more opportunities for the impacts of one patient to be offset by the arrival behavior of other patients. For these reasons, overbooking is particularly advantageous in larger clinics that see more patients during a clinic session.

Generalized results

The results of our simulation studies demonstrate that the benefit of overbooking varies with a clinic's utility structure and size. To provide more general results that can be applied to a wide range of clinical settings, we undertook regression analyses of our experimental results with the goal of developing functional representations of both the benefits and costs of overbooking.

Three separate regression studies were conducted using the 50 factor combinations of independent variables N , S , and R , with mean patient wait time, mean clinic overtime, and mean productivity as dependent variables. Functional representations were found for expected patient wait time $\bar{W}(N, S)$, expected clinic overtime $\bar{O}(N, S)$, and expected productivity $\bar{P}(N, R)$:

$$\bar{W}(N, S) = 2.608S + .067N - .061SN - 2.653S^2, \quad (5)$$

$$\bar{O}(N, S) = 4.692S + .109N - .104SN - 4.658S^2, \quad (6)$$

$$\bar{P}(N, R) = 1 - .357R + .003RN + .143R^2. \quad (7)$$

For each of these regression equations, all coefficients were significant at .95 confidence levels, and each had coefficients of determination (R^2) greater than .98.

Using these results, expected overbooking utility $\bar{U}_N(N, S)$ can be determined using equations (3), (5), and (6) with appropriate values for parameters π , ω , and τ :

$$\bar{U}_N(N, S) = \pi[\bar{A} - SN] - \omega\bar{W}(N, S) - \tau\bar{O}(N, S). \quad (8)$$

Note that the expected number of arriving patients \bar{A} is simply the clinic size N by construction of K , the number of appointments scheduled in a clinic session.

To use these results, the staff and administration of the previously described mental health clinic ($N = 16$, $S = 70\%$, $\pi = 1$, $\omega = .5$, and $\tau = 1.2$) applied equations (5)–(8) to determine that overbooking would significantly increase both patient access and provider productivity. With overbooking, the clinic would expect to see 16 patients a day compared to 11.2 patients without overbooking, an increase in patient access of 43%. Implementing an overbooking policy would increase provider utilization from 70% to 92%, for an increase in effective productivity of 32%. The trade-off for these benefits is increased patient waiting time and increased clinic overtime. Expected patient waiting time is equal to $.91D$ and expected clinic overtime is $1.58D$, where D is the duration of an appointment.

To balance the benefits of overbooking with its costs, the clinic applied utility function (8) and found that overbooking would increase net clinic utility by 2.45 units. This positive result was interpreted by the clinic as the net equivalent of providing service to almost two and one-half additional patients each day for each of its clinicians. As a public health provider, the clinic viewed overbooking as an attractive means to quickly, easily, and economically enhance its service to the community through improved utilization of its existing resources. The impact of this policy is that new patients, who otherwise would be turned away due to lack of available clinic capacity, can receive the mental health care services they need. When patients are denied access to outpatient services, they are often forced to use providers of last resort (such as hospitals) at a much higher cost to the community. With increased outpatient access, all of the clinic's patients receive more timely follow-up appointments, improving the quality of their care and their satisfaction with services.

SENSITIVITY ANALYSIS AND IMPLEMENTATION

We now extend our examination of appointment overbooking to examine its sensitivity to stochastic service times, and then discuss several issues relevant to implementing our results in practice.

Sensitivity to Service Time Uncertainty

Our analysis of overbooking is predicated on several assumptions, most important, that of deterministic service times. To examine the impact of overbooking when service times are uncertain, we conducted simulation experiments, in addition to those described in the appointment overbooking model section, for three clinic size levels $N \in \{10, 30, 50\}$ and three levels for show rate $S \in \{.1, .5, .9\}$. *Provider service time variability* was defined as the coefficient of variation of service time, $c_s = \sigma/\mu$, where σ is the standard deviation of service time and μ is its mean.

Table 2: Effect of service time variability on clinic performance with overbooking.

Clinic Size N	10	10	10	30	30	30	50	50	50	
Show Rate S	90%	50%	10%	90%	50%	10%	90%	50%	10%	
Average Patient Waiting Time W										
	.00	.30	.83	1.17	.63	1.68	2.34	1.11	2.27	3.14
	.35	.49	.93	1.24	1.02	1.88	2.47	1.61	2.53	3.35
	.58	.69	1.05	1.33	1.44	2.16	2.70	2.20	2.93	3.64
	.71	.82	1.14	1.40	1.72	2.37	2.87	2.55	3.21	3.89
	1.00	1.09	1.33	1.57	2.26	2.76	3.22	3.28	3.76	4.35
Average Clinic Overtime O										
	.00	.53	1.52	2.20	1.02	2.79	3.95	1.84	3.71	5.17
	.35	.84	1.68	2.32	1.64	3.10	4.17	2.62	4.12	5.47
	.58	1.17	1.92	2.48	2.32	3.57	4.55	3.54	4.78	5.93
	.71	1.40	2.08	2.61	2.79	3.94	4.84	4.13	5.18	6.33
	1.00	1.86	2.40	2.94	3.58	4.57	5.39	5.27	6.10	7.08
Average Server Productivity P (utilization)										
	.00	.940	.862	.808	.958	.911	.879	.972	.928	.904
	.35	.910	.849	.799	.938	.901	.873	.957	.920	.899
	.58	.879	.828	.785	.917	.889	.863	.940	.911	.891
	.71	.861	.816	.774	.904	.880	.855	.930	.903	.884
	1.00	.819	.782	.745	.874	.854	.834	.904	.880	.865
Average Net Overbooking Utility U_N ($\pi = 1.0, \omega = .5, \tau = 1.2$)										
	.00	.11	2.76	5.77	1.17	10.81	21.09	2.65	19.42	37.23
	.35	-.35	2.53	5.60	.23	10.35	20.76	1.46	18.79	36.77
	.58	-.85	2.18	5.36	-.81	9.64	20.19	.05	17.80	36.07
	.71	-1.19	1.94	5.17	-1.51	9.08	19.76	-.84	17.18	35.46
	1.00	-1.87	1.46	4.69	-2.72	8.14	18.93	-2.56	15.80	34.33

Variable service times were modeled using a Gamma distribution with parameter pairs (α, β) of (8.0, .125), (3.0, .333), (2.0, .5), and (1.0, 1.0), providing a total of five levels of service time variability $c_s \in \{0.0, .35, .58, .71, 1.0\}$, including the original experiments where $c_s = 0$. The Gamma distribution was used, because it is bounded from below by 0 (no negative service times) and is skewed to the right with a long right tail allowing the possibility of extended service times, perhaps corresponding to emergency service. Summary results are reported in Table 4, which shows average patient waiting time W , average clinic overtime O , average provider productivity P (utilization), and the average net overbooking utility benefit U_N gained with overbooking.

The results of these follow-on experiments indicate that clinic performance deteriorates with increasing service time variability. Average patient wait times increase, clinic overtime increases, and provider productivity decreases. These results are consistent with many studies showing that variability and uncertainty is antithetical to efficient operations (Hopp & Spearman, 2001). Table 4, however, shows that, while overbooking utility benefits U_N attenuate with increasing service time variability, appointment overbooking can continue to provide positive benefits even when service time variability is very high ($c_s = 1.0$). In cases in which

overbooking provides significant utility benefits (large clinics with frequent no-shows), overbooking continues to provide significant benefits as service variability increases. This result should provide assurance to practitioners that appointment overbooking can bring valuable benefits to clinic performance even when provider service times are highly uncertain. In cases in which the benefits of overbooking are marginal (smaller clinics with low no-show rates), the benefits of overbooking remain marginal or can become negative as service time uncertainty increases. This latter result indicates that caution should be exercised when considering overbooking in smaller clinics with low no-show rates and in which service times are highly uncertain.

Implementation Issues

The results of our model can be applied easily in practice by using regression equations (5) and (6) to determine expected patient wait time $\bar{W}(N, S)$ and clinic overtime $\bar{O}(N, S)$, and then applying utility equation (3) to determine whether overbooking will provide positive net utility U_N . If expected net utility is negative, then overbooking is not advised; if positive, then overbooking will be beneficial. However, a second assumption of our model is that patient arrivals can be scheduled in continuous time with a time interval between appointments of $T = N/K$. This could easily lead to appointments being scheduled at odd and inconvenient intervals (every 18 minutes, for example), depending on the values of N and K . In practice, appointments are normally scheduled at multiples of 5, 10, 15, 20, 30, 45, or 60 minutes, depending on the expected length of the appointment and other circumstances. For example, if $T = 18$ minutes, then appointment start times could be rounded up or down to the nearest five-minute increment. Note that the appointment start times are rounded, not the interval T between appointments.

Rounding down would favor reduced provider idle time at the expense of increased patient wait time. Rounding up would result in increased provider idle time and clinic overtime, but with the benefit of reduced patient wait time. The choice of these trade-offs will depend on the preferences of the clinic as manifested by its selection of cost parameters ω and τ . If patient waiting costs ω are greater than clinic overtime costs τ ($\omega > \tau$), then rounding up appointment start times might be advised. Conversely, if $\omega < \tau$, then rounding down would be preferred.

CONCLUSIONS AND FUTURE RESEARCH

In this article we have evaluated the utility of appointment overbooking as a means to mitigate the negative impact of patient no-shows in health care clinics, where no-shows reduce provider productivity and clinic efficiency, increase health care costs, and limit patient access to necessary services. While there is extensive literature that investigates overbooking in transportation services, appointment overbooking has not been previously studied as a possible remedy for the negative impacts of no-shows in health care clinics or other appointment-driven service operations. As the first to formally study appointment overbooking in this context, we believe that this article makes several significant contributions to the health care

management literature, and to the improvement of clinic operations and decision-making practices in health care services.

First, we showed how scheduling complexity increases when appointment overbooking is used to compensate for no-shows. By demonstrating the dynamics of patient arrival uncertainty in both the timing and the number of no-shows, we differentiated clinic overbooking from overbooking for revenue management in transportation services. We also showed that patient wait time and provider overtime inevitably increase with appointment overbooking, so clinic administrators must be willing to endure increased patient waiting times and increased clinic overtime if appointment overbooking is used.

Second, we introduced a new utility model to evaluate appointment overbooking in terms of trade-offs between the benefits of serving additional patients and the costs of increased patient wait time and provider overtime. By adjusting subjective weights for the relative importance of patient access, patient wait times, and clinic overtime, our utility model provides health care service providers and administrators with an easy and effective means to determine whether or not overbooking is an attractive policy for a specific clinic.

Third, with simulation experiments, regression analysis, and sensitivity experiments, we showed that appointment overbooking can significantly improve clinic performance by increasing patient access and improving clinic productivity, which translates into reduced clinic costs and improved patient satisfaction and outcomes. Like revenue management in transportation operations, we showed that appropriate appointment overbooking can provide an effective and important means to improve capacity utilization and increase patient access to health care services. In an era of rapidly escalating health care costs and calls for improved health care efficiency, these are indeed attractive benefits. While it would be incorrect to conclude that all clinics should always overbook, our results indicate that patient overbooking can often lead to substantially improved net clinic performance. Follow-on sensitivity experiments indicated that appointment overbooking can often improve clinic utility even when service time variability is large.

Fourth, we provided managerial insight into the practical use of appointment overbooking in real-world clinics and demonstrated its application in a large, publicly-funded mental health clinic. We identified situations in which overbooking is likely to be beneficial and in which it is likely to be counterproductive. All else equal, overbooking provides more utility when clinics serve larger numbers of patients, no-show rates are higher, and service variability is lower. Also, waiting times and overtime are disproportionately greater for small clinics than for large ones. With overbooking, provider productivity declines as no-show rates increase, but at a rate that is much lower than without overbooking, and overbooking can provide robust productivity performance results even in the presence of large no-show rates. Even with highly variable service times, many clinics will benefit from overbooking, but caution should be used in small clinics with low no-show rates and highly variable services times.

A perhaps surprising result of our research is that overbooking provides more utility when no-show rates are high because of the increased opportunity to overbook, and thereby boosts expected productivity. At high show rates (which

is equivalent to low no-show rates), the probability of patients showing up is by definition higher. Thus, the risks of overtime and patient wait time are higher if overbooking is employed, thereby incurring greater costs that might exceed the benefits expected from seeing more patients. Therefore, for sufficiently high costs of patient wait time and overtime, there is more benefit in overbooking when no-show rates are higher, and higher costs incurred by overbooking when no-show rates are lower.

A number of promising avenues exist for extending clinic overbooking research. First, in this article we set appointment start times to intervals of $T = N/K$, but other appointment scheduling policies might better balance patient waiting costs and clinic overtime costs with the benefits of increased patient accessibility. An interesting line of research would be to study the effectiveness of various scheduling policies in combination with overbooking such as double-booking (Rohleder & Klassen, 2002) and wave scheduling (Chung, 2002). When double-booking, a clinic schedules two appointments to start at the same time, and when wave-scheduling, uses cyclical booking patterns to improve clinic performance. These might prove to be effective means of implementing an overbooking policy.

Second, in this article we fixed the number of scheduled appointments to $K = N/S$ rounded to the nearest integer. It would be useful to investigate changes in clinic utility using different overbooking policies that vary the number of patients overbooked. And, finally, in some clinics patients do not see a specified provider but can see any one of several providers that are available, requiring a more complex overbooking model.

Each of these avenues suggests potential extensions to the research reported in this article. The impact of overbooking on clinic costs, capacity, working conditions, and customer service, which we examined in this article, highlights the importance of continued research to determine how the complexity of overbooked appointment schedules can be managed to maximize its utility. [Received: March 2006. Accepted: February 2007.]

REFERENCES

- Bailey, N. T. (1952). A study of queues and appointment systems in hospital outpatient departments, with special reference to waiting-times. *Journal of the Royal Statistical Society, Series B, 14*(2), 185–199.
- Barnhart, C., Belobaba, P., & Odoni, A. R. (2003). Applications of operations research in the air transport industry. *Transportation Science, 37*(4), 368–391.
- Baum, N. H. (2001). Control your scheduling to ensure patient satisfaction. *Urology Times, 29*(3), 38–43.
- Barron, W. M. (1980). Failed appointments: Who misses them, why they are missed, and what can be done. *Primary Care, 7*(4), 563–574.
- Bean, A. B., & Talaga, J. (1995). Predicting appointment breaking. *Journal of Health Care Marketing, 15*(1), 29–34.

- Blanco White, M. J., & Pike, M. C. (1964). Appointment systems in out-patients' clinics and effect of patients' unpunctuality. *Medical Care*, 2(3), 133–145.
- Brahimi, M., & Worthington, D. J. (1991). Queueing models for out-patient appointment systems—A case study. *Journal of the Operational Research Society*, 42(9), 733–746.
- Campbell, J. D., Chez, R. A., Queen, T., Barcelo, A., & Patron, E. (2000). The no-show rate in a high-risk obstetric clinic. *Journal of Women's Health & Gender-Based Medicine*, 9(8), 891–895.
- Cayirli, T., & Veral, E. (2003). Outpatient scheduling in health care: A review of literature. *Production and Operations Management*, 12(4), 519–549.
- Chesanow, N. (1996). Can't stay on schedule? Here's a solution. *Medical Economics*, 73(21), 174–180.
- Chung, M. K. (2002). Tuning up your patient schedule. *Family Practice Management*, 9(1), 41–48.
- Cole, R. M. (2003). Four steps to a streamlined schedule: Analyzing your schedule in minute detail can help you gain control and keep your practice at peak performance. *Review of Optometry*, 140(6), 33–35.
- Colorado Department of Human Services, Division of Mental Health. (2005). MHSIP consumer survey: Consumers' perceptions of care in the community mental health system in Colorado, accessed March 1, 2006, available at http://www.cdhs.state.co.us/dmh/PDFs/de_MHSIP_MHCtech_rpt_fy05.pdf.
- Easton, F. F., & Goodale, J. C. (2005). Schedule recovery: Unplanned absences in service operations. *Decision Sciences*, 36(3), 459–448.
- Fetter, R. B., & Thompson, J. D. (1966). Patients' waiting time and doctors' idle time in the outpatient setting. *Health Services Research*, 1(1), 66–90.
- Garuda, S. R., Javalgi, R. G., & Talluri, V. S. (1998). Tackling no-show behavior: A market-driven approach. *Health Marketing Quarterly*, 5(4), 25–45.
- Grunebaum, M., Luber, P., Callahan, M., Leon, A. C., Olfson, M., & Portera, L. (1996). Predictors of missed appointments for psychiatric consultations in a primary care clinic. *Psychiatric Services*, 47(8), 848–852.
- Ho, C., & Lau, H. (1992). Minimizing total cost in scheduling outpatient appointments. *Management Science*, 38(12), 1750–1763.
- Hopp, W., & Spearman, M. (2001). *Factory physics* (2nd ed.). New York: McGraw-Hill.
- Kimes, S. E. (1989). Yield management: A tool for capacity-constrained service firms. *Journal of Operations Management*, 8(4), 348–363.
- Klassen, K. J., & Rohleder, T. R. (1996). Scheduling outpatient appointments in a dynamic environment. *Journal of Operations Management*, 14(2), 83–101.

- Kruse, G. R., & Rohland, B. M. (2002). Factors associated with attendance at a first appointment after discharge from a psychiatric hospital. *Psychiatric Services, 53*(4), 473–476.
- Larkin, H. (1999). Satisfaction pays. *American Medical News, 42*(30), 17.
- Lieberman, W. (2004). Revenue management trends and opportunities. *Journal of Revenue and Pricing Management, 4*(1), 91–99.
- Lieberman, W. (2005). How times have changed for the revenue management professional! *Journal of Revenue and Pricing Management, 4*(20), 109–110.
- McCarthy, E. L. (2002). Physician office productivity improvement through operations analysis and process redesign. *The Journal of Ambulatory Care Management, 25*(4), 37–52.
- McCarthy, K., McGee, H. M., & O'Boyle, C. A. (2000). Outpatient clinic waiting times and non-attendance as indicators of quality. *Psychology, Health & Medicine, 5*(3), 287–293.
- McGill, J. I., & Van Ryzin, G. J. (1999). Revenue management: Research overview and prospects. *Transportation Science, 33*(2), 233–256.
- Metters, R., & Vargas, V. (1999). Yield management for the nonprofit sector. *Journal of Service Research, 1*(3), 215–226.
- Murray, M., & Berwick, D. M. (2003). Advanced access: Reducing waiting and delays in primary care. *Journal of the American Medical Association, 289*(8), 1035–1040.
- Rohleder, T. R., & Klassen, K. J. (2002). Rolling horizon appointment scheduling: A simulation study. *Health Care Management Science, 5*(3), 201–209.
- Rust, C. T., Gallups, N. H., Clark, W. S., Jones, D. S., & Wilcox, W. D. (1995). Patient appointment failures in pediatric resident continuity clinics. *Archives of Pediatrics & Adolescent Medicine, 149*(6), 693–695.
- Sharp, D. J., & Hamilton, W. (2001). Non-attendance at general practices and outpatient clinics: Local systems are needed to address local problems. *British Medical Journal, 323*(7321), 1081–1082.
- Shonick, W., & Klein, B. W. (1977). An approach to reducing the adverse effects of broken appointments in primary care systems. *Medical Care, 15*(5), 419–429.
- Sweeney, D. R. (1996). Your office: A lot of things will have to change. *Medical Economics, 73*(7), 97–102.
- Toh, R. S., & Raven, P. (2003). Perishable asset revenue management: Integrated Internet marketing strategies for the airlines. *Transportation Journal, 42*(4), 30–44.
- Tonges, M. C. (1985). Quality with economy: Doing the right things for less. *Nursing Economics, 3*(4), 205–211.
- Van Ryzin, G. J., & Talluri, K. T. (2003). Revenue management. In R. W. Hall (Ed.), *Handbook of transportation science*. Boston: Kluwer Academic Publishers, 599–659.

Visser, J. (1979). Selecting a suitable appointment system in an outpatient setting. *Medical Care*, 17(12), 1207–1220.

Visser, J., & Wijngaard, J. (1979). The outpatient appointment system: Design of a simulation study. *European Journal of Operational Research*, 3(6), 459–463.

Warden, J. (1995). 4.5 Million outpatients miss appointments. *British Medical Journal*, 310(6), 1158.

Weatherford, L. R., & Bodily, S. E. (1992). A taxonomy and research overview of perishable-asset revenue management. *Operations Research*, 40(5), 831–844.

Welch, J. D., & Bailey, N. T. (1952). Appointment systems in hospital outpatient departments. *The Lancet*, 1(22), 1105–1108.

Wright, P. D., Bretthauer, K. M., & Côté, M. J. (2006). Reexamining the nurse scheduling problem: Staffing ratios and nursing shortages. *Decision Sciences*, 37(1), 39–70.

APPENDIX

In this appendix, we develop closed-form clinic performance measures for $N \in \{1, 2, 3\}$ and $K \in \{1, 2, 3\}$. It is convenient to focus on the number of patients scheduled K , rather than on the clinic size N . For $K = 2$, feasible clinic sizes are either $N = 1$ with $.40 < S \leq .667$, or $N = 2$ with $.80 < S \leq 1$. Denoting ‘1’ as a patient who arrives for a scheduled appointment and ‘0’ for a no-show patient, the four possible permutations of patient behavior and associated performance characteristics are shown in the following table:

Show–No-show Sequence	Probability of Sequence	Accumulated Wait Time	Wait Time per Patient	Session Finish Time
1,1	S^2	$D - T$	$(D - T)/2$	$2D$
1,0	$S(1 - S)$	0	0	D
0,1	$(1 - S)S$	0	0	$T + D$
0,0	$(1 - S)^2$	0	0	0

Within the appropriate ranges of S indicated above, the expected number of arriving patients, expected wait time for arriving patients, and expected clinic finish time are

$$\begin{aligned} \bar{A} &= 2S^2 + S(1 - S) + (1 - S)S = 2S \\ \bar{W} &= S^2(D - T)/2 \\ \bar{F} &= S^2(2D) + S(1 - S)D + S(1 - S)(T + D) = S(TR + 2D). \end{aligned} \tag{A1}$$

For the special case $S = 1$, the only possible show and no-show sequence is represented by the first row, and then, because $T = SD = D$, there is no patient wait time.

When $K = 3$ and $.167 \leq S < .5$, imputed clinic size is $N = 1$; $N = 2$ for $.5 \leq S < .833$, and $N = 3$ for $.833 \leq S \leq 1$. There exist nine permutations of patient arrivals with $K = 3$ as shown in the following table:

Show–No-show Sequence	Probability of Sequence	Accumulated Wait Time	Wait Time per Patient	Session Finish Time
1,1,1	S^3	$3(D - T)$	$(D - T)$	$3D$
1,1,0	$S^2(1 - S)$	$D - T$	$(D - T)/2$	$2D$
1,0,1	$S^2(1 - S)$	$\text{Max}\{0, D - 2T\}$	$\text{Max}\{0, D - 2T\}/2$	$\text{Max}\{2T, D\} + D$
1,0,0	$S(1 - S)^2$	0	0	D
0,1,1	$S^2(1 - S)$	$D - T$	$(D - T)/2$	$T + 2D$
0,1,0	$S(1 - S)^2$	0	0	$T + D$
0,0,1	$S(1 - S)^2$	0	0	$2T + D$
0,0,0	$(1 - S)^3$	0	0	0

Therefore, for the ranges of S indicated above, the expected number of arriving patients, expected wait time for arriving patients, and expected clinic finish time are:

$$\begin{aligned} \bar{A} &= 3S \\ \bar{W} &= S^2[(D - T) + (1 - S) \text{max}\{0, D - 2T\}/2] \\ \bar{F} &= 3S^3 + S^2(1 - S)(5D + \text{max}\{2T, D\} + T) + 3S(1 - S)^2(T + D). \end{aligned} \tag{A2}$$

From these results, the expected net utilities \bar{U}_N shown in Table 3 can be derived using equation (3).

Linda R. LaGanga is the director of quality systems at the Mental Health Center of Denver. She received her PhD in business administration with a major in operations research and a minor in strategy and entrepreneurship from the University of Colorado at Boulder. She received her MS in operations research and statistics from Rensselaer Polytechnic Institute, an MA in clinical mental health counseling from Rivier College, and her BS in applied mathematics and computer science from the University of Rhode Island. She has held senior management positions in the software and health care industries and is a licensed professional counselor who has provided direct clinical services and supervised other clinicians for licensure. Her research is motivated by her interest in developing systems that promote effective resource allocation in health care and other service organizations to expand service capacity, particularly to increase underserved populations’ access to services.

Stephen R. Lawrence is an associate professor of operations management in the Leeds School of Business of the University of Colorado at Boulder. He received BS and MS degrees in engineering from Purdue University and his doctorate in industrial administration from Carnegie-Mellon University. Professor Lawrence’s research interests include service operations, supply chain management, technology management, and topics in renewable energy. His research has been published

in a number of well-known academic journals, and he is the recipient of a number of teaching awards. Professor Lawrence also has held several administrative posts at the Leeds School including faculty director of executive programs, associate dean for programs, and interim dean. Prior to his academic career, Professor Lawrence worked in industry for eight years, culminating in five years as vice president for operations for a metal casting plant in the Midwest.