Managing No Shows Via Appointment System Design

A THESIS
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Christine Tupy

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

DIWAKAR GUPTA

April, 2011

© Christine Tupy 2011
ALL RIGHTS RESERVED

Acknowledgements

There are many people that have earned my gratitude for their contribution to my time in graduate school.

First and foremost, I would like to thank my supervisor, Dr. Diwakar Gupta, who offered continuous support. I am grateful for your invaluable assistance and immense knowledge. Without your patience and guidance this project would not have been possible. I could not have imagined having a better adviser and mentor for my M.S. study.

I would also like to thank those that were on my examination committee, Dr. Francois Sainfort and Dr. Sandra Potthoff. Without your insights, time, and cooperation this project would not have been successful.

For their efforts and assistance, a special thanks as well to my fellow SCORLAB colleagues.

To each of the above, I extend my deepest appreciation.

Dedication

This thesis is dedicated to my parents who have offered me unconditional love and support since the beginning of my studies.

Abstract

Healthcare systems struggle to match clinic availability with patient demand. Balancing supply with demand becomes more complicated when patients cancel their appointments late or do not show up for their appointments. Late cancellations and no shows can disrupt the delivery of care, increase patient wait times, and lower clinic revenues. These consequences together make it difficult for healthcare systems to contain costs while continuing to improve the quality of care. In this thesis, we study the appointment system design problem with overbooking and consider a new strategy, clinic configuration, to manage no shows. In particular, we evaluate two different clinic configurations – one in which patients with higher no show rates are streamed to special clinics and the other in which all patients are randomly assigned to available doctor clinics. We formulate models to determine which strategy is best suited for a clinic and present numerical experiments which demonstrate that clinic configuration can have a significant affect on the expected net revenue in some cases.

Contents

List of Figures

Chapter 1

Introduction

In 2008, US health care expenditures accounted for approximately 16.2% of the nation's gross domestic product, equivalently more than $2.3 trillion (Kimbuende et al. 2010). This amount is more than three times the $714 billion spent in 1990, and more than eight times the $253 billion spent in 1980 (Kimbuende et al. 2010). Many factors, such as medical technology advancements, greater prevalence of chronic illnesses, and the aging of the population, have contributed to the rapid growth in health care spending. The spending growth has been accompanied by increasing pressure on health service providers to contain costs while continuing to improve the quality of care.

This thesis is concerned about outpatient clinic capacity, which is expensive to make available, and whose utilization is affected by patient no shows and late cancellations. Outpatient care, a form of medical care where patients are not required to stay overnight, includes both primary and specialty care visits. Primary care is patients' recurring point of contact with health systems, which requires coordination with other service providers. Specialty care consults, which usually occur via physician referrals, are specialized medical services provided to patients by health care specialists. This thesis focuses on comparing two different clinic configurations in the primary care setting – one in which patients with higher no show rates are streamed to special clinics and the other in which all patients are randomly assigned to available doctor clinics.

No shows are common in health care settings. In a survey of 200 pediatric resident continuity clinics, no shows rates were found to range from values as small as 3% to as much as 80% (Rust et al. 1995). In a survey of family practice residency clinics (i.e. clinics that train medical residents), Hixon et al. (1999) found that 60.5% of the clinics that responded had a reported no show rate less than 21%, 35.3% had a reported no show rate between 21%-50%, and a small percentage (1.4%) of clinics had a reported rate of greater than 50%.

No shows and late cancellations leave unfilled open appointment slots in providers' schedules and limit patients' access to their care providers. High no show rates cause loss of revenue, longer patient wait times, poor quality of care, and these factors together serve to increase medical costs. No shows also cause high levels of dissatisfaction among health care providers and patients. Ho and Lau (1992) found that no shows were a major factor among three environmental factors – no shows, service time variation, and the number of patients per clinic session – that affects providers' utilization. Multiple issues have been found to be correlated with high no show rates including demographic and socioeconomic status, situational factors, appointment lead times, age, gender, and history of no shows. Reasons for no shows typically vary significantly from one health care setting to another making it difficult to develop a universal strategy that can prevent or limit their negative consequences.

In addition to no shows, there are other factors that affect a clinic's ability to run efficiently, manage capacity and allow patients to be seen when they want. These factors include the random arrival and service processes, and patient/provider preferences (Gupta and Denton 2008). The vast majority of patients access outpatient clinics' services through a scheduled appointment. However, in addition to scheduled appointments, clinics also serve the needs of those that walk-in and urgent cases. These unpredictable cases make the design of an appointment system challenging because it is hard to anticipate how many appointment slots will be needed for each type of appointment. The problem is further exacerbated by, service time variability, which makes it difficult to decide how much time to reserve for each service.

Patients generally prefer to see their doctor of choice at a convenient time on a certain day. These preferences change from one encounter request to another, as well as over time. For example, depending on the medical urgency, a patient may sometimes not care which physician provides the service for some appointment requests, whereas the same patient may be very particular about his/her physician choice for some other requests. Physicians may also place restrictions on how available slots may be filled and how many overbook slots to allow. These patient/provider preferences make matching supply and demand more complicated. A well-designed appointment system streamlines the delivery of outpatient care, which can help improve utilization of providers' capacity, and reduce crowding in clinics and overall costs of care while honoring patients' and providers' preferences.

The focus of this thesis is on a comparison of two different models of clinic design. It uses clinics' expected net revenue as the measure of effectiveness of each design. Net revenue is not to be confused with profit motive. The approach assumes that the clinic earns a reward for serving each patient and incurs costs for turning away patients as well as for the overtime work done by service providers. Thus, it incorporates patients' and providers' viewpoints as well. Moreover, net revenue does not account for the clinic's fixed cost of making providers' services available to patients. By choosing the relative values of the three parameters mentioned above, a clinic can gain deeper understanding of the parameter ranges within which one design dominates the other. Our models allow random arrival process, ability to estimate patient no show rates by cohort, and multiple physicians. We do not model service time variability because we consider decisions that arise after a clinic divides its service providers' time into fixed intervals, called appointment slots. In reality, service times are random and service providers try to remain on schedule by varying pace of work. The problem of setting optimal appointment lengths with random service times is an important problem in its own right and has been discussed in several other papers, see e.g. Denton and Gupta (2003). We also do not model patient and provider preferences. Such problems have also been addressed elsewhere; see for example, Wang and Gupta (2011) who develop a framework for the design of the next generation of appointment systems that dynamically learn and update patients' preferences, and use this information to improve booking decisions.

In particular, this study is motivated by an analysis of appointments data from two Veterans Health Administration (VHA) facilities. It was found that for both sites approximately 12% to 15% of appointments are not utilized due to patient no shows and late cancellations. These results are consistent with the findings from the VA Office of Inspector General, who estimated that the VHA wastes 18.9% of its clinic appointment capacity representing $76M annually in wasted resources. It was also discovered that no show rates varied by patient group, e.g. different primary care no show rates were observed among veterans who also had appointments with mental health clinics. The purpose of this study is to evaluate the impact of clinic configuration on the management of no shows.

Literature on no shows falls into one of two categories: papers that focus on evaluating appointment booking strategies (including overbooking) upon assuming that no show rates are exogenous, and papers that evaluate strategies for reducing no show rates. For example, clinics may book more appointments than available capacity either by double or triple booking in certain slots, or reducing the length of each slot in an attempt to increase provider capacity utilization. However, overbooking may also cause too many patients to arrive during a clinic session, which could potentially increase patients' in-clinic wait times, increase providers' overtime, and increase variability of provider workloads from day to day. In contrast, positive effects of overbooking include improve patient access and provider productivity if implemented carefully. Within the second category of papers, researchers first identify the reasons for no shows, such as patient apathy, provider availability, or language barriers, and eliminate them. For example, increasing the number of appointments patients have with their preferred providers (also known as PCPs) may lead to lower no show rates. But, such strategies may also exacerbate provider workload imbalances, increase patient wait times, and reduce capacity utilization. In summary, various approaches to manage no shows result in positive as well as negative consequences and models as well as experimental validation are needed to evaluate which combination of strategies are likely to produce the best results.

Papers published in the open literature have not considered clinic configuration as a possible strategy to manage no shows. Our model evaluates two different clinic configurations. Each clinic configuration differs in the way patients are assigned to a clinic or provider. We allow optimal overbooking within each configuration and determine if one method is more efficient than the other over different ranges of clinic parameter values. The two configurations, refereed to as the mixed and the specialized strategies, are shown in a schematic below.

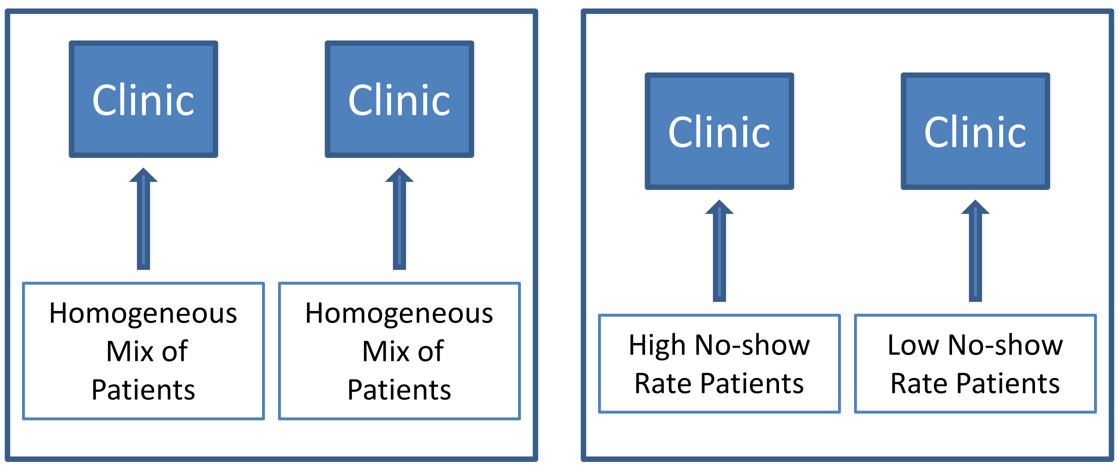

Figure 1.1: Appointment System Design: Mixed Strategy (left panel) and Specialized Strategy (right panel).

Figure 1.1 shows a clinic with two individual providers and different approaches for streaming patients into the clinics. For the mixed strategy, each provider faces a homogeneous mix of patients. In contrast, for the specialized strategy, patients are streamed based on their no show rates; one physician receives high no show rate patients and the other receives low no show rate patients. This can be achieved by designating a specific clinic or physician for the high no show rate patients.

In this paper we develop a model to calculate the expected net revenue of a clinic which includes the revenue generated minus the cost of turning away demand and the cost of provider overtime, to determine what clinic configuration strategy is most effective. We find the optimal, i.e. net revenue maximizing, booking limit for a clinic given the fact that some patients may not show up for their appointments. We assume that all arriving patients either get scheduled for an appointment or get turned away and told to find an alternate date/clinic if the booking limit is reached. We also investigate the effect of the parameters, revenue, cost of turned away demand, and cost of overtime, on the expected net revenue function. Lastly, we examine the two clinic configuration strategies further using a multiple day approach where patients are booked into one of several future days according to a strategy that maximizes total expected revenue.

The rest of this thesis is organized as follows. In the next section, we give a brief literature review and discuss various aspects of the problem of appointment system design that have been explored and modeled before. Then in, Section 3, we formulate models used to determine which strategy is best suited for a clinic. Section 4 contains a discussion about how to estimate the parameters used in the models. Section 5 contains the results of this investigation. Finally, in Section 6, we conclude the thesis and discuss avenues for further research.

Chapter 2

Literature Review

In biomedical literature, many papers focus on outpatient scheduling and on identifying features of appointment systems that help health care systems match demand with capacity in a manner that utilizes resources more efficiently. In this chapter, we describe the modeling techniques that have been used before, and explain how our model relates to them. In particular, we include an analysis of previous works in this chapter that investigate the effect of the number of doctors, service time variability, no shows, and incentives. The literature on no shows and late cancellations can be classified into one of the following two categories: (1) papers that study the use of overbooking to mitigate the negative effects of no shows, and (2) papers that study intervention strategies to reduce no shows. We discuss each stream of literature next.

Many papers model clinics as single-server systems for the purpose of appointment system design (Cayirli and Veral 2003). This is because in primary care setting, each patient picks a preferred provider and belongs to that physician's panel. Patients prefer to be matched with their preferred doctors when medical needs arise. This practice results in better care for the patient, better understanding of the patient as a person, and better patient-physician communication (Gonnella et al. 1980). Therefore, for a clinic system involving multiple physicians, we assume independent arrival streams for each doctor.

In appointment scheduling literature a variety of service time distributions are used. Some papers use deterministic (e.g. LaGanga and Lawrence 2007, and Robinson and Chen 2010), while others use random service times (e.g. Patrick et al. 2008, and Robinson and Chen 2003). The coefficient of variation, which is the standard deviation divided by the mean, is a commonly used measure of the variability of consultation times (Cayirli and Veral 2003). Denton and Gupta (2003) find that optimal solutions do depend on skewness. In Kim and Giachetti (2006), variance in patient service time is relatively not that significant, and, therefore, the total time in a regular session may be approximated with the mean service time of each appointment multiplied by the total number of appointments. In our model, we assume that all appointment slots are given the same length of time (constant). If a patient requires more time, he or she may be scheduled for more than one appointment slot in succession.

Cayirli and Veral (2003) contains a review of the literature on appointment scheduling for outpatient services, covering 80 papers that span a period of over 50 years. Only a few of these papers consider the impact of no shows. We investigated the appointment scheduling literature that has appeared since 2003 in which there is a possibility that a patient may not show up for his or her appointment. Several papers in the literature develop analytical and simulation models that assume all patients have the same no show probability (e.g. Kaandorp and Koole 2007, LaGanga and

Lawrence 2007, Hassin and Mendel 2008, and Robinson and Chen 2010). Kim and Giachetti (2006) uses probability distributions for both no shows and walk-ins. Other studies have categorized patients into no show groups depending on their attributes. If patients belong to a group, say j, then they are endowed with a probability $p_j > 0$ of arriving and a probability $1 - p_j$ of not showing (e.g. Chakraborty et al. 2009, Muthuraman and Lawley 2008, Zeng et al. 2010). Some studies take into account the fact that no show probability depends on appointment delays, the time between when a patient calls and the date of the appointment (e.g. Qu et al. 2007, and Liu et al. 2010). Green and Savin (2008) consider that the probability of a no show is non-decreasing in the size of the appointment queue at the time when the patient makes an appointment. In our model, we divide patients into two show rate categories, high and low.

There are many different performance measures that are used to evaluate appointment systems and there is currently no approach that incorporates every single measure. Most literature that uses modeling techniques to improve clinic appointment systems focuses on minimizing patient wait times, physician idle times, and physician overtime (e.g. Kaandorp and Koole 2007, Chakraborty et al. 2009, Robinson and Chen 2010, and Hassin and Mendel 2008), and some focus on matching demand with capacity (e.g. Qu et al. 2007). Some other measures include expected revenue or profit (e.g. LaGanga and Lawrence 2007, Kim and Giachetti 2006, Muthuraman and Lawley 2008, and Zeng et al. 2010), expected patient backlog and the probability of getting a same-day appointment (e.g. Green and Savin 2008). In our formulation, we evaluate a clinic configuration based on its expected revenue, net of the cost of turning away demand and physician overtime.

As mentioned earlier, the first category of papers study the use of overbooking to mitigate the negative effects of no shows. Kim and Giachetti (2006) develop a stochastic overbooking model for determining the optimal number of patient appointments to accept to maximize expected total profits for diverse healthcare environments. They conclude that the stochastic mathematical overbooking model can significantly improve expected total profits for healthcare providers, compared with the naive statistical overbooking approach that simply adds the mean number of no shows minus the mean number of walk-ins to the number of appointments to accept.

LaGanga and Lawrence (2007) introduce a 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. They determine that overbooking provides greater utility when clinics serve larger numbers of patients, no show rates are higher, and service variability is lower, and that even with highly variable service times, many clinics will achieve positive net results with overbooking. Muthuraman and Lawley (2008) develop an overbooking process that accommodates the detailed requirements and dynamics of the clinical scheduling environment and utilize a model to predict the probabilities that patients will not show up. They make the assumption that estimates of the patients' no show probability can be made based on patient attributes and historical data for the patient or for the group of patients with similar attributes. This is similar to our formulation, where we also assume that patients can be categorized into groups based on no show rates. In Zeng et al. (2010), the authors observe that homogeneous overbooking models using the mean value of show-up probabilities are not enough to build high quality schedules. The variance of no show probabilities have a significant impact on the performance of overbooked schedules.

Another stream of papers focus on strategies that can be used to reduce no shows. These strategies include, service, communication, price, and location. Service strategies include methods such as minimizing the time to appointment and maintaining patient satisfaction (e.g. Benjamin-Bauman et al. 1984). A patient's feeling of ease and the thoroughness of the physician's examination can affect the patient's appointment keeping behavior (Pearce 1979). Communication strategies include appointment reminder systems (mail, telephone, or physician reminders) (e.g. Cohen et al. 1980, and Koren et al. 1994) and patient education (e.g. Bigby et al. 1983, and Guse et al. 2003). Examples of price strategies include using a reward system for showing (e.g. Rice and Lutzker 1984), charging a fee for no shows, and conducting follow-up appointments by phone. Because traveling to the clinic entails extra cost to the patient, it seems logical that conducting some appointments by telephone may reduce no shows (Bean 1992). Lastly, reducing transportation is a location strategy. Vikander et al. (1986) found that transportation difficulties were the most frequently cited barrier to appointment keeping. Our investigation is not focused on reducing no shows. Our goal is to determine if a different clinic configuration strategy can help reduce the negative effect of no shows on a clinic's revenues.

Chapter 3

Models

Typically clinics have multiple physicians and multiple types of patients. For example, patient types can be categorized into either new or existing. New patients are those that are scheduling an appointment for the very first time or patients that have been referred to a physician specialist for consultation. Existing patients are those that have had an appointment before and have been seen by that physician earlier. Primary care physicians' panels typically evolve over time as new patients join the panel and existing patients leave. With a mixed clinic strategy patients choose which physicians they would like to enroll with, while a specialized clinic strategy streams patients by assigning them to a specific physician. Health systems can also create specialized clinic days by designating certain days when each physician sees new patients. This effectively creates two types of clinics for each physician with potentially different no show rates. A question that arises is which strategy is better when booking limits are chosen optimally within each strategy.

3.1 Revenue Function

We begin with a single server model of a single day of operation of a clinic in which there is a single average show rate for all patients. Later we generalize this model to consider multiple physicians and multiple days of operation.

Patients arrive punctually to their scheduled appointments and all appointments are completed during their scheduled times. These assumptions are made to keep mathematical models tractable. They may be relaxed in detailed simulation models of clinic operations after promising strategies are identified by analytical models. Overbooking is allowed. To find the optimal overbooking limit one must first determine a clinic's net revenue as a function of the number of patients overbooked and then find the best booking limit such that the revenue is maximized. Therefore we begin by formulating the revenue function. For this purpose, the following notation is used, with upper case notation representing a random variable.

k = capacity (number of appointment slots)

b = the booking limit

r = revenue received per patient seen

π = the cost of turning away demand

c = the cost of overtime

X = demand

x = a realization of demand

$\lambda = \mathrm{E}(X)$, the average demand

p = the show rate of all patients

$Y(m)$ = the number of patients that show up given that m are booked

It is assumed that $Y(m) = \sum_{i=1}^{m} Z^{(\mathrm{i})}$, where $Z^{(\mathrm{i})}$ are independent identically distributed Bernoulli random variables with parameter p. $Y(m)$ therefore follows a binomial distribution with parameters m and p, where m is the number of trials and p is the show rate. In numerical examples, we assume that X follows a Poisson distribution with parameter λ which is equal to the expected demand for a single day. However, no distributional assumption is required for the analysis presented in this section. The notation $(\cdot)^+$ stands for max($\cdot$,0) and $a \wedge b$ for min(a,b).

The revenue generated upon using a booking limit b and when the demand realization is x is denoted as $\rho(b, x)$. This can be written as the following equation:

$$\rho(b, x) = rE[Y(b \wedge x)] - \pi(x - b)^+ - cE[(Y(b \wedge x) - k)^+]. \tag{3.1}$$

The revenue function $\rho(b, x)$ can be simplified to the following:

$$\rho(b, x) = \begin{cases} rE[Y(x)] - cE[(Y(x) - k)^+], & \text{if } x \leq b, \\ rE[Y(b)] - \pi(x - b) - cE[(Y(b) - k)^+], & \text{if } x > b. \end{cases}$$

The terms on the right hand side of $\rho(b, x)$ are the expressions for the two possible cases that may occur. Either the demand could be larger than the booking limit, or the demand could be smaller than or equal to the booking limit. If the demand is larger than the booking limit some patients will be turned away, but if the demand is smaller than or equal to the booking limit then all patients would get scheduled. So $\rho(b) = E_X[\rho(b, X)]$ is equal to:

$$= \sum_{x=0}^{\infty} \rho(b, x) P(X = x)$$

$$
\begin{aligned}
&= \sum_{x=0}^{b}(rpx - c\sum_{i=0}^{x}(i-k)^{+}P(Y(x)=i))P(X=x) \\
&\quad + \sum_{x=b+1}^{\infty}(rbp - \pi(x-b) - c\sum_{i=0}^{b}(i-k)^{+}P(Y(b)=i))P(X=x) \\
&= rp\sum_{x=0}^{b}xP(X=x) - c\sum_{x=0}^{b}(\sum_{i=0}^{x}(i-k)^{+}P(Y(x)=i))P(X=x) \\
&\quad + rpbP(X>b) + \pi bP(X>b) - \pi\sum_{x=b+1}^{\infty}xP(X=x) \\
&\quad - c\sum_{x=b+1}^{\infty}(\sum_{i=0}^{b}(i-k)^{+}P(Y(b)=i))P(X=x) \\
&= rp\sum_{x=0}^{b}xP(X=x) + (rp+\pi)bP(X>b) \\
&\quad - c\sum_{x=k}^{b}(\sum_{i=k}^{x}(i-k)P(Y(x)=i))P(X=x) \\
&\quad - cP(X>b)\sum_{i=k}^{b}(i-k)P(Y(b)=i) - \pi(E(X) - \sum_{x=0}^{b}xP(X=x)). \qquad (3.2)
\end{aligned}
$$

To obtain the last equality in (3.2), we use the fact that

$$
\sum_{x=0}^{b}\sum_{i=0}^{x}(i-k)^{+} = \begin{cases} 0, & \text{if } x < k \text{ or } i < k, \text{ and} \\ \sum_{x=k}^{b}\sum_{i=k}^{x}(i-k), & \text{otherwise.} \end{cases}
$$

Using Equation (3.2) one can calculate the expected net revenue for any combination of parameters $(k, r, \pi, c, \lambda, p, b)$. The revenue function was coded in MATLAB 7.7.0 to obtain numerical solutions. Next, we show how to find the optimal booking limit b.

3.2 Optimal Booking Limit

We need to establish sufficient conditions under which $\rho(b)$ has monotone decreasing differences in order to determine the revenue maximizing booking limit. Consider the terms on the right hand side of $\rho(b) - \rho(b-1)$ after conditioning upon the values of X and then taking pathwise differences.

First Term:

$$rE[Y(b \wedge X) - Y((b-1) \wedge X)] = \begin{cases} rE[Y(X) - Y(X)], & \text{if } X \leq b-1 \text{ and,} \\ rE[Y(b) - Y(b-1)], & \text{if } X \geq b. \end{cases}$$

Therefore the above difference is equal to $rE[Y(b) - Y(b-1)]P(X \geq b)$, because $Y(m)$ and X are independent.

Second Term:

$$-\pi E[(X-b)^+ - (X-(b-1))^+] =$$

$$\begin{cases} -\pi E[(X-b) - (X-(b-1))] = \pi, & \text{if } X \geq b, \\ 0, & \text{otherwise.} \end{cases}$$

Therefore the second term is equal to $\pi P(X \geq b)$.

Third Term: Given that $k \leq b$, we obtain

$$-cE[(Y(b \wedge X) - k)^+ - Y((b-1) \wedge X) - k)^+] =$$

$$\begin{cases} -cE[Y(X) - k)^+ - (Y(X) - k)^+] = 0, & \text{if } X \leq b-1, \\ -cE[Y(b) - k)^+ - (Y(b-1) - k)^+], & \text{if } X \geq b. \end{cases}$$

Now we must consider

$$E[(Y(b) - k)^+ - (Y(b-1) - k)^+] = \begin{cases} 0, & \text{if } Y(b-1) < k, \\ E[Y(b) - Y(b-1)], & \text{if } Y(b-1) \geq k. \end{cases}$$

Therefore the third term is equal to $-cE[(Y(b)-Y(b-1))\times \mathbf{1}_{(Y(b-1)\geq k)}]P(X\geq b)$. By combining all of the terms and plugging in $Y(m)=\sum_{i=1}^{m} Z^{(\mathrm{i})}$, we get the following:

$$\begin{aligned}
\rho(b)-\rho(b-1) &= P(X\geq b)[rE[\sum_{i=1}^{b} Z^{(i)}-\sum_{i=1}^{b-1} Z^{(i)}]+\pi \\
&\quad -cE[(\sum_{i=1}^{b} Z^{(i)}-\sum_{i=1}^{b-1} Z^{(i)})|\sum_{i=1}^{b-1} Z^{(i)}\geq k]P(\sum_{i=1}^{b-1} Z^{(i)}\geq k)] \\
&= P(X\geq b)[rE(Z^{(b)})+\pi-cE(Z^{(b)})P(\sum_{i=1}^{b-1} Z^{(i)}\geq k)] \\
&= P(X\geq b)[rp+\pi-cpP(\sum_{i=1}^{b-1} Z^{(i)}\geq k)] \qquad (3.3)
\end{aligned}$$

In the last equality above, we use the fact that $E(Z^{(i)})=p$ for every i. Therefore, the difference of the differences, $[\rho(b+1)-\rho(b)]-[\rho(b)-\rho(b-1)]$ is equal to:

$$\begin{aligned}
&= P(X\geq b+1)[rp+\pi-cpP(\sum_{i=1}^{b} Z^{(\mathrm{i})}\geq k)] \\
&\quad -P(X\geq b)[rp+\pi-cpP(\sum_{i=1}^{b-1} Z^{(\mathrm{i})}\geq k)] \\
&\leq P(X\geq b+1)[rp+\pi-cpP(\sum_{i=1}^{b} Z^{(\mathrm{i})}\geq k)] \\
&\quad -P(X\geq b)[rp+\pi-cpP(\sum_{i=1}^{b} Z^{(\mathrm{i})}\geq k)] \\
&= -P(X=b)[rp+\pi-cpP(\sum_{i=1}^{b} Z^{(\mathrm{i})}\geq k)] \qquad (3.4)
\end{aligned}$$

Equation (3.4) <0 if the expression $rp+\pi-cpP(\sum_{i=1}^{b} Z^{(\mathrm{i})}\geq k)>0$. This is implied by the following sufficient condition $rp+\pi-cp\geq 0$. However, that leads to a very high booking limit because if $rp+\pi-cp\geq 0$, then $rp+\pi-cpP(\sum_{i=1}^{b} Z^{(\mathrm{i})}\geq k)\geq rp+\pi-cp\geq 0$, and $\rho(b+1)-\rho(b)>0$ for every b in the support of the distribution of X. Therefore, we assume that $rp + \pi - cp < 0$. We also know that $rp+\pi-cpP(\sum_{i=1}^{0} Z^{(\mathrm{i})} \geq k) = rp+\pi > 0$ and that the term $rp+\pi-cpP(\sum_{i=1}^{b} Z^{(\mathrm{i})} \geq k)$ is decreasing in b. Therefore, it follows from the intermediate value theorem that there exists a $\bar{b}$ such that $rp+\pi-cpP(\sum_{i=1}^{m} Z^{(\mathrm{i})} \geq k) < 0$ for every $m > \bar{b}$ and the inequality is reversed when $m \leq \bar{b}$.

Suppose that $\rho(m) - \rho(m-1) < 0$ for all $m > \bar{b}$. Then, the optimal overbooking limit lies in $[k, \bar{b}]$. Moreover, because $rp+\pi-cpP(\sum_{i=1}^{b} Z^{(\mathrm{i})} \geq k) \geq 0$ when $b \in [k, \bar{b}]$, clearly, $\rho(b) - \rho(b-1)$ is monotone decreasing in b over $[k, \bar{b}]$ and for $b > \bar{b}$, $\rho(b)$ is decreasing in b. Hence, the optimal booking limit is either $\bar{b}$ or an interior solution to the first order optimality equation.

In summary, the optimal booking limits for different values of show rate (p) are found by finding the smallest value of m such that

$$rp + \pi - cpP(\sum_{i=1}^{m} Z^{(\mathrm{i})} \geq k) < 0. \tag{3.5}$$

$P(\sum_{i=1}^{m} Z^{(\mathrm{i})} \geq k)$ follows a binomial distribution with parameters m and p. Using equation (3.5) one can calculate the optimal booking limit given values (k, r, π, c, p). The optimality criterion was coded in MATLAB 7.7.0 to obtain numerical solutions.

3.3 Problem Formulation

In the previous section, we developed a method to calculate the revenue function and the optimal booking limit for a single physician system. Now, the question that we want to answer is whether a mixed or a specialized strategy performs better. We do so by comparing the expected revenue under each strategy when booking limits are chosen optimally. First, we describe the model set-up for each strategy.

Consider two clinics that each serve some demand from a population of P patients. Within this population, patients are categorized by a show rate of either high (p_1) or low (p_2). For the specialized strategy we pick λ_1 and λ_2 for the two clinics. For the mixed strategy, each clinic serves a homogeneous mix of patients each with average show rate (p), which equals $(\lambda_1 p_1 + \lambda_2 p_2)/(\lambda_1 + \lambda_2)$. Each clinic has the same capacity and each has the same booking limit, which is determined from the criterion in (3.5).

In the specialized strategy, the population of P patients is streamed into two groups corresponding to high (p_1) and low (p_2) show rates. One clinic serves the high show rate patients, while the other clinic serves the low show rate patients. Each clinic has the same capacity, but each has a different booking limit, which is determined from the criterion in (3.5). We set appointment request rates in the two clinics such that the offered load is the same under mixed and specialized strategies.

3.3.1 A Comparison of Strategies

To compare the mixed strategy to the specialized strategy, a MATLAB program was developed that can determine which strategy gives the better expected net revenue for different values of high and low show rates. Offered loads are set equal to clinic capacities in all scenarios. For varying values of p, p_1, and p_2 the program calculates arrival rates $\lambda = k/p$, $\lambda_1 = k/p_1$, and $\lambda_2 = k/p_2$ so that the expected load of patients is balanced for each case. Next, the program finds the 3 optimal booking limits b, b_1, and b_2 for input parameters (k, r, π, c, p_1, p_2). Then the program calculates the expected net revenue per unit time for each case. That is, it gives the expected net revenue for each group of parameters: $(r, \lambda, b, p, \pi, c, k)$,

$(r, \lambda_1, b_1, p_1, \pi, c, k)$, and $(r, \lambda_2, b_2, p_2, \pi, c, k)$. Then the program determines the best strategy by comparing $2\times$ revenue$(r, \lambda, b, p, \pi, c, k)$ with revenue$(r, \lambda_1, b_1, p_1, \pi, c, k)$ + revenue$(r, \lambda_2, b_2, p_2, \pi, c, k)$. We show examples of these calculations in the Results section.

3.4 Multiple Physicians

A variant of the above approach was developed to test what happens when there are multiple physicians. In particular, for the specialized case, N_1 physicians are selected to serve high no show rate (p_1) patients and N_2 physicians are selected to serve low no show rate (p_2) patients. Then for the mixed case, $N = N_1 + N_2$ and all physicians serve both patient types. The program that compares these two cases can be executed in one of two ways. The first way is to have the physicians each have their own individual arrival stream of patients. The second way is to pool the physicians arrival streams together. If the physician arrival streams are pooled, then the mixed strategy is always superior because capacity pooling dominates any benefits that come from the ability to set different booking limits for different types of patients. But in that case patients may not be matched with their preferred providers. Therefore, in the next chapter, we consider examples in which physicians have individual patient request arrival streams.

3.5 Multiple Booking Days

To investigate the two strategies further, a multiple day computational approach was developed where patients are booked into slots within a booking horizon according to a cost comparison. Each day a random number of patients arrive and are assigned one by one as they arrive to scheduled slots. The cost of overbooking a patient to a specific slot is obtained and compared with the cost of scheduling the patient on a future day. Another aspect of this model is that patients are either booked into one of 0, 1, ..., D days or turned away, where D represents the booking horizon. With a fixed value of D, it is possible that maximum capacity maybe reached and patient demand may need to be turned away. That is, beyond a certain level of congestion additional patients may be told to call later or try alternative clinics. However, when capacity is adequate and the corresponding value of D is sufficiently large, turn-aways are not very likely. We use the following additional notation in this section.

k = capacity

p = the show rate

c = the overtime cost

π = delay cost per day of scheduling a patient in the future

D = the finite limit on how far in the future patients may be booked

t = day index, how many days in the future from the current period, i.e. $t \in (0, 1, 2, ..., D)$

$\vec{S}$ = the vector of the current schedule = $[S_0, S_1, ..., S_D]$ where S_0 is the number of patients schedule for day 0

$Y(z)$ = the number of patients that show up on an arbitrary day given z were booked

$V(z)$ = the expected overtime cost of booking z patients

X = demand, where $P(X = x)$ has a truncated Poisson distribution

The number of patients that arrive is known and every day after scheduling we observe some overtime cost with expectation $V(S_0)$ and some delay cost $\sum_{t=1}^{D} S_t \pi t$. There is a penalty for not serving a patient right away and there is an additional cost of overtime if capacity is exceeded by scheduling the patient on a certain day. Every time a patient is booked we accrue some revenue, and possible overtime and delay cost. When a patient is scheduled t days into the future, the cost of scheduling delay equals πt, whereas when the patient is the zth patient to be overbooked, the cost of overtime equals $V(z+1) - V(z)$, where $V(z) = cE[(Y(z) - k)^+]$. Figure 3.1 shows

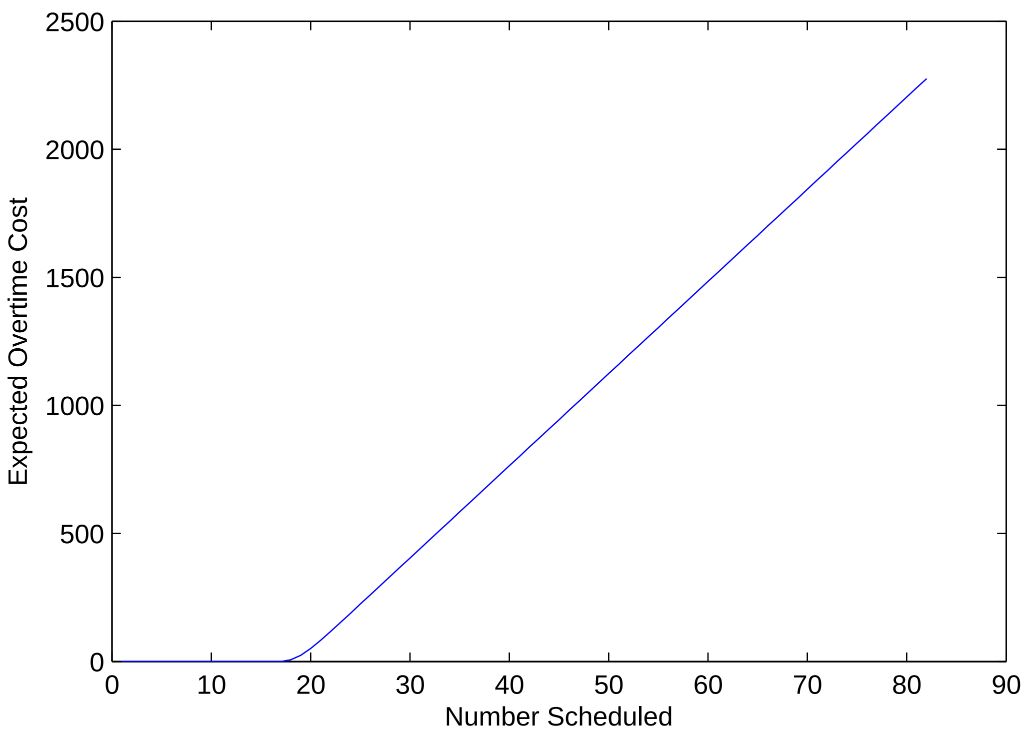

Figure 3.1: Overtime Plot

the plot of $V(z)$. Patients are scheduled one by one to the day slot with the cheapest associated cost of assignment. The change in cost due to scheduling a patient to some slot on day t is given below:

$$\Delta C_t = V(S_t + 1) - V(S_t) + \pi t \tag{3.6}$$

Let $\Delta\vec{C}$ represent the cost vector of scheduling a patient to a certain slot, i.e. $\Delta\vec{C} = (\Delta C_0, \Delta C_1, ..., \Delta C_D)$.

The daily demand in this model is assumed to be a truncated Poisson distribution. It is important to chose the maximum value of demand ($\bar{x}$) to be sufficiently large so that the probability of the truncation at $\bar{x}$ is low. That is, there must be a finite value of $\bar{x}$ to ensure that the state space is finite and numerical analysis is tractable. The vector $\vec{S}$ can be thought of as a function that maps Q, the total number of patients in the schedule to each future clinic day. To mark this relationship, we also use notation $\vec{S}(Q)$. So, we may focus first on finding steady state numbers of patients in the schedule and then use $\vec{S}(Q)$ to find the total expected overtime and total expected delay costs. Therefore, we focus on finding the distribution of Q next.

We know that Q is equal to the total number of patients in the schedule $= \sum_{t=0}^{D} S_t$. Let there be some function $F(Q)$ that returns the number of patients scheduled to be served on day t=0 given that Q are in the system, i.e. $F(Q) = S_0$. The plot of $F(Q)$ is shown in Figure 3.2.

First, we want to find $\bar{q}$, which is the maximum number of patients that can be in the system at once. To find $\bar{q}$, we calculate $F(Q)$ for each Q and when $F(Q) = \bar{x}$ we label that value of Q as $\bar{q}$. It is easy to see that $F(Q) \leq F(Q+1) \leq F(Q) + 1$ for each $Q \geq 0$. That is, $0 \leq F(Q+1) - F(Q) \leq 1$ and $F(Q)$ is increasing in Q at a rate less than the rate of increase of Q. This means that by increasing Q one at a

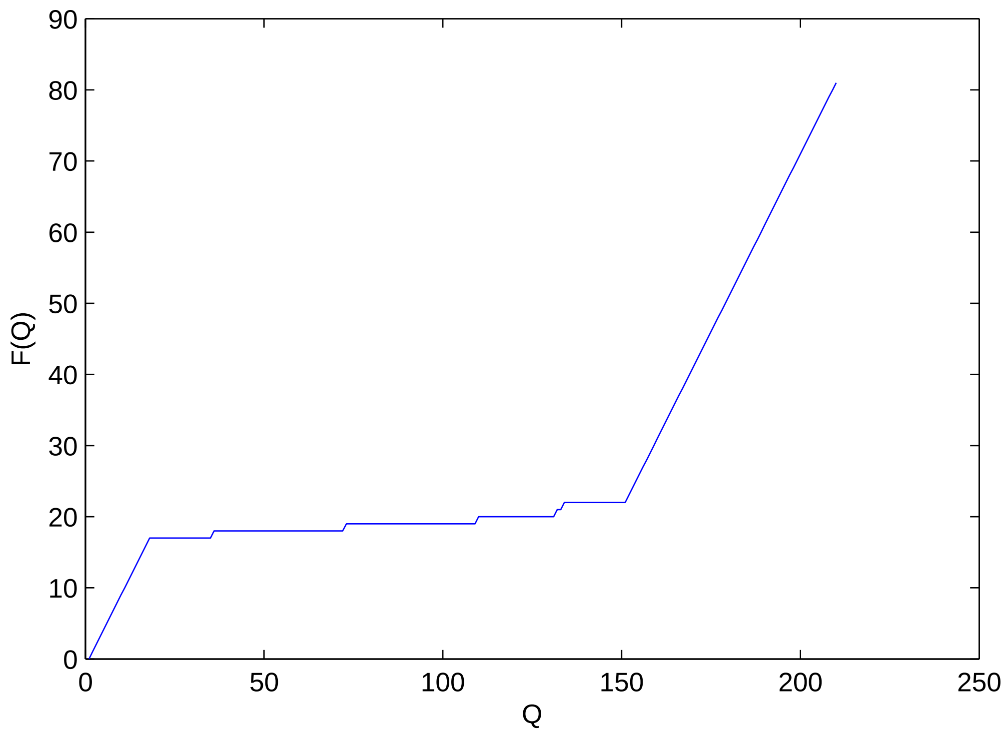

Figure 3.2: Number Served Given Q Are In The System

time, it would be possible to find a value of Q such that $F(Q) = Q$. This value of Q is labeled $\bar{q}$.

Next we find $\Delta\vec{C}$ for all values of $Q = 0, \ldots, \bar{q}$, and update $\vec{S}$ by adding each patient one by one to $\vec{S}$ at S_t if that assignment minimizes the value of ΔC_t for all t. To find the steady state for this system we calculated the transition probabilities, which we call matrix $A_{i,j}$ with dimension $\bar{q} \times \bar{q}$, i.e. $A_{i,j}$ is the probability of ending up with j patients in the system given we started with i. To find each element in A, let Q^- and Q^+ denote queue lengths (patients not yet scheduled) at the beginning and end of an arbitrary day. Then

$$\begin{aligned} A_{i,j} &= P(Q^+ = j | Q^- = i) \\ &= \begin{cases} \sum_{l=0}^{\bar{x}} P(X = l) \times \mathbf{1}_{(F(i+l)=i+l-j))}, & \text{if } i \geq j, \\ \sum_{l=j-i}^{\bar{x}} P(X = l) \times \mathbf{1}_{(F(i+l)=i+l-j))}, & \text{if } i < j. \end{cases} \end{aligned} \tag{3.7}$$

We define α as the steady state vector of probabilities for the number of patients in the system. Then we solve the matrix equation $A^T\alpha = \alpha$, where α has dimensions $[0, \bar{q}]$, to get the steady state probability vector α. We can use the steady state probabilities to find the expected overtime and expected delay cost for this system.

$$E[OT] = \sum_{j=0}^{\bar{q}-\bar{x}} \sum_{l=0}^{\bar{x}} V(F(j+l)) \times \alpha(j) \times P(X = l) \tag{3.8}$$

$$E[W] = \pi \sum_{j=0}^{\bar{q}-\bar{x}} j \times \alpha(j) \tag{3.9}$$

where $\alpha(j)$ is the probability of ending in state j regardless of the state we started in. The values of $E[OT]$ and $E[W]$ can be investigated and compared for the two strategies described in Figure 1.1. A MATLAB program was developed to compare the two strategies for this model setup and the results of this investigation are shown in the next chapter.

Chapter 4

Parameter Estimation

In order for clinic managers to use the models we have presented to test which clinic configuration strategy is best suited for a clinic, they need to estimate the parameters used in the models. In this section we discuss how clinic mangers may estimate those parameters. The parameters are r (revenue), π (cost of turning away demand), c (cost of overtime), p (show rate), and X (demand).

To calculate the revenue parameter first determine the revenue per patient and obtain that value for each patient seen. It is important that the revenue be the actual dollar amount that was realized from the amount billed for that visit. For all appointments, the average of the realized revenue received per patient visit should be designated as r.

To calculate the cost of turning away demand two important things must be considered. The first is if a patient is not booked for an appointment they may never come back, and the second is if a patient is not booked when they want an appointment and schedule an appointment further into the future, then they have to wait for that appointment. This is undesirable because the patient's health could deteriorate over time, which could require more expensive care. It could also be dissatisfying to the patients because they have to wait for their appointment, which could in turn lead to patients enrolling with other service providers. Therefore, the cost of turning away demand should depend on the revenue that was not received, the patient wait costs, and operating costs. The fraction of revenue loss will come from appointments that do not get booked because patients who are turned away do not book on a future date. A few papers have addressed the estimation of the cost of patient wait time. Russel (2009) values patients' time according to its opportunity cost. This paper uses the economic reasoning that hourly wage rate can represent the opportunity cost of time. Robinson and Chen (2010) provide queuing-based bounds on the value of relative patient wait cost that are fairly tight and are easily calculated from two readily observable parameters: the average number of patients waiting and the doctor's utilization. Lastly, in Yabroff et al. (2005) researchers have identified categories of key medical services for colorectal cancer care and then estimated patient time associated with each service category using data from national surveys. The average service frequencies were combined with estimates of patient time for each category of service, and the value of patient time was assigned. The operating costs would be equal to the cost of providing care to a patient. So, the cost of turning away demand should be equal to average revenue per patient times the fraction of patients who do not book on a future date minus the operating costs plus the cost of patient wait time.

To estimate the cost of overtime, the clinic manager would need to estimate the hourly cost of overtime wages for staff and providers for each day. Then, an average needs to be calculated over all days of operation.

There are several methods that a clinic manager can use to estimate show rates. One option is to calculate the overall average show rate for all patients, but this is not very realistic since some patients may no show more than others. Another option is to classify the patients into multiple show rate categories. This can be done by analyzing a patients' history of no shows. Also, the patients could be grouped into different categories based on the type of care they need (i.e., mental health, primary care) and each category may have its own show rate.

To estimate demand, a clinic could use a couple different types of data. The number of appointment requests that come in each day is the demand. If the desired date, the day when a patient wants to be seen or the day the physician recommended the patient to be seen, is known then the demand for any day would be the number of patients that had that day as their desired date. If the desired date data is not available then the demand can be estimated by using past appointment data from a recent period of time, for example 12 months. Find the number of appointments that occurred in that 12 month period and the number of working days. Then divide the number of appointments by the number of working days.

Chapter 5

Results

In this section we demonstrate some numerical results we obtain when comparing the mixed strategy to the specialized strategy. First we show the results of both strategies when there are only two providers/clinics. Then we show the results for multiple providers. Next we show how unit revenue, cost of turning away a patient, and physician overtime affect expected net revenue by varying each value one at a time. Lastly, we show the results of experiments involving multiple booking days.

5.1 Mixed vs. Specialized Strategies

A MATLAB program was developed that can determine which strategy gives rise to a higher net revenue for different values of high and low show rates. Outlined below are the input parameters that were used to compare the two strategies:

p_1 = 0.8, 0.8025, 0.8050, ..., 1, this means it starts at 0.8 and increments by 0.0025 up to 1

p_2 = 0.8, 0.8025, 0.8050, ..., 1

$k = 16$

$r = 30$

$\pi = 5$

$c = 40$

The show rates p_1 and p_2 were chosen to be in the 0.8 to 1 range to account for typical show rates seen in health care clinics. The capacity was assigned to be 16, assuming that a physician has an 8 hour day with appointments having a service time of 30 minutes each, which means the physician can see a total of 16 patients each day. The parameters r, π, and c were chosen arbitrarily considering the constraint $rE(Z) + \pi - cE(Z) < 0$, where $E(Z) =$ the show rate. Later in this section we investigate the consequences of changing r, π, and c on the clinic's expected net revenue.

Figure 5.1 shows the results for the above input parameters. If mixed is better than specialized a (red) dot is plotted on a graph at (p_1, p_2). If specialized is better than mixed a (blue) ”x” is plotted on the graph at (p_1, p_2). And if the two strategies expected net revenues are equal then a (black) ”+” is plotted on the graph at (p_1, p_2). A strategy is better than the other if the expected revenue from equation (3.2) is larger. Figure 5.1 shows that when there are only two physicians in a clinic, overall a specialized strategy is better when the two show rates differ significantly. However, when the two show rates are not too different and the booking limits are similar, then a mixed strategy is preferred.

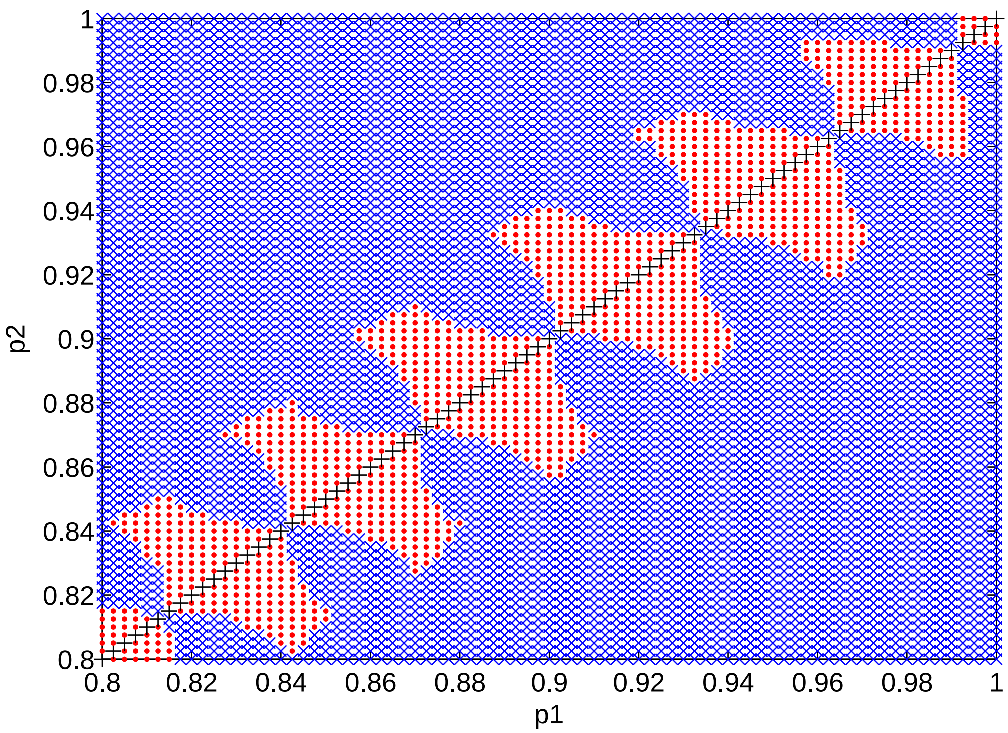

Figure 5.1: Mixed vs. Specialized - Two physicians

5.2 Multiple Physicians

In this section we show some numerical results obtained when comparing the mixed strategy to the specialized strategy for different numbers of physicians. The idea here is to model a clinic with multiple physicians in which only one physician is assigned to the low show rate patients while all other physicians are assigned to the high show rate patients. Below are the input parameters used to compare the mixed strategy to the specialized strategy for this setup:

$p_1 = 0.8, 0.8025, 0.8050, ..., 1$, this means it starts at 0.8 and increments by 0.0025 up to 1

$p_2 = 0.8, 0.8025, 0.8050, ..., 1$

$k = 16$

$r = 30$

$\pi = 5$

$c = 40$

$N_1 = 9$

$N_2 = 1$

Figure 5.2 shows mixed versus specialized for physicians having individual arrival

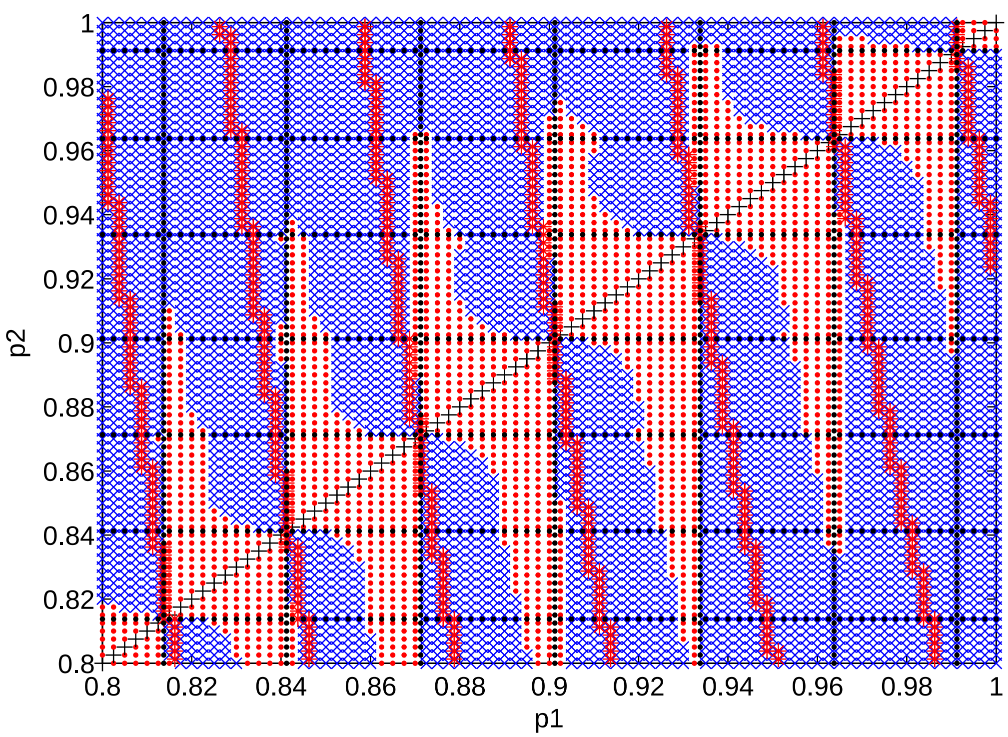

Figure 5.2: Mixed vs. Specialized With $N_1 = 9$ and $N_2 = 1$

streams where $N_1 = 9$ and $N_2 = 1$. Also included in Figure 5.2 are the boundary lines where the booking limits b, b_1, and b_2 change. Again, if mixed is better than specialized a (red) dot is plotted on a graph at (p_1, p_2). If specialized is better than mixed a (blue) "x" is plotted on the graph at (p_1, p_2). And if the two strategies revenues are equal then a (black) "+" is plotted on the graph at (p_1, p_2). The black dots signify changes in b_1 and b_2, and the red stars signify changes in b.

Figure 5.2 shows that when the values of p_1 and p_2 are close together and the values for b, b_1, and b_2 are the same, then a mixed strategy is best. However, once the values of p_1 and p_2 start to differ significantly and the values for b, b_1, and b_2 also differ, then the best strategy depends on the parameters. Lastly, when the values of p_1 and p_2 differ greatly then the majority of the time specialized strategy is better, but the best strategy still depends on the values of all of the parameters.

5.3 Parametric Comparisons

Next, we change the values of r, π, and c to investigate how these parameters affect the difference in expected net revenue between the mixed and specialized strategies. To perform this investigation we define a parameter β as follows:

$$\beta = (rp + \pi)/cp. \tag{5.1}$$

Because $rp+\pi-cp < 0$, each variable r, π, and c has an effect on β. While varying p_1 and p_2 from 0.8 to 1, we can vary one of the variables (r, π, c) while keeping the other two constant, which allows us to vary β from 0 to 1. We know that as the revenue, r, increases so does β, and as the cost of delay, π, increases so does β, but as the cost of overtime, c, increases β decreases asymptotically. Therefore, since those behaviors are already known, we decided to investigate the percentage of expected net revenue that would be lost if we were to pick mixed clinics where specialized clinics were better when β ranges from 0 to 1. This was done to determine if the difference in expected net revenue between the two strategies is significant. Each variable (r, π, c) is programmed to vary such that β will go from 0 to 1 and is plotted against the maximum percent difference in expected net revenue for varying values of p_1 and p_2.

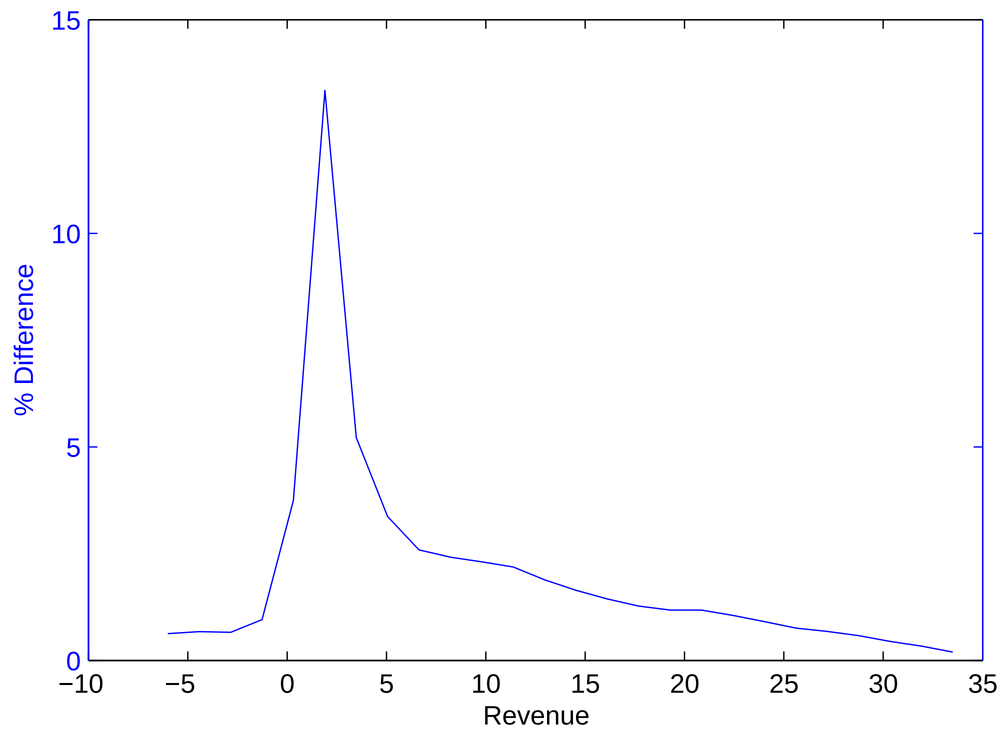

Figure 5.3: Revenue vs. Maximum Percent Difference in Profit

Figures 5.3, 5.4, and 5.5 show us that most of the differences in expected net revenue percentage are small if we are to choose mixed clinics where specialized was

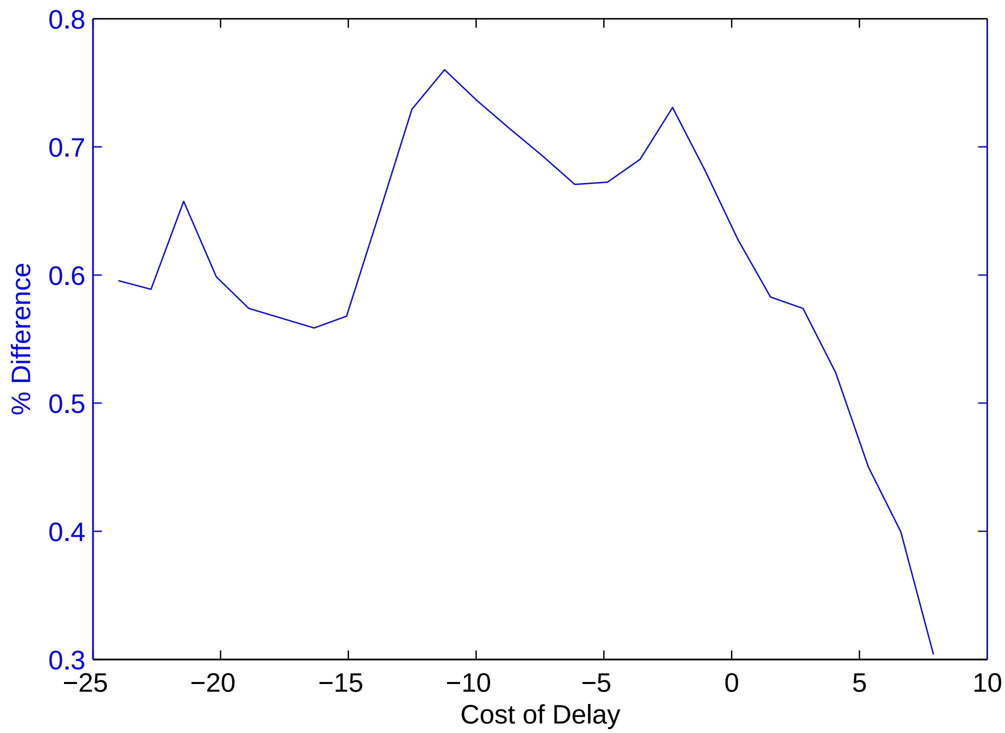

Figure 5.4: Delay Cost vs. Maximum Percent Difference in Profit

better. However, in a few cases, such as when r is varied, we get a maximum difference in expected net revenue of 13.34% which is quite significant. The revenue value for parameters $(r = 1.9, \lambda = 17.758, b = 16, p = .901, \pi = 5, c = 40, k = 16)$ is 12.48, for parameters $(r = 1.9, \lambda_1 = 19.51, b_1 = 18, p_1 = 0.82, \pi = 5, c = 40, k = 16)$ is 8.85, and for parameters $(r = 1.9, \lambda_2 = 16, b_2 = 16, p_2 = 1, \pi = 5, c = 40, k = 16)$ is 19.45. The reason for this large difference in expected net revenue percentage could be due to the fact that in the pooling (mixed) case the booking limit is set to not allow for any overbooking. Therefore, if a patient does not show up for a scheduled appointment, in the mixed case there is no chance that the slot gets filled since no overbooking was

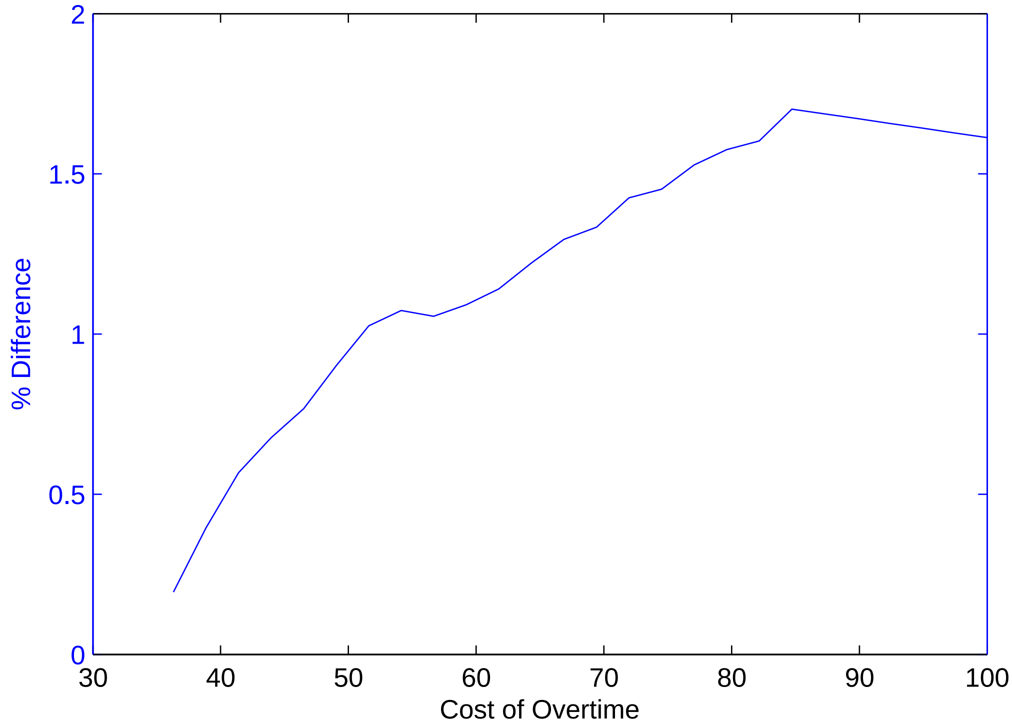

Figure 5.5: Overtime Cost vs. Maximum Percent Difference in Profit

allowed.

5.4 Multiple Booking Days

In this section we show the numerical results obtained for the multiple booking days problem when comparing the mixed clinic strategy to the specialized clinic strategy. The MATLAB program developed to investigate this model finds the expected cost that is accrued each day for both strategies and then compares them. The expected cost is equal to the expected overtime cost plus the expected delay cost. Below are the input parameters used to compare the mixed strategy to the specialized strategy for this setup:

$p_1 = 0.8, 0.81, 0.82, ..., 1$, this means it starts at 0.8 and increments by 0.1 up to 1

$p_2 = 0.8, 0.81, 0.82, ..., 1$

$k = 16$

$\pi = 5$

$c = 40$

$X = [0, 40]$

The maximum value for demand was chosen to be sufficiently large enough such that the probability of truncation is low at the maximum value of X. As explained earlier, there must be a finite value of the maximum value of X to ensure that the state space is finite and numerical analysis is tractable. In this case, we chose the maximum value of X to be 2.5 $\times k$. Figure 5.6 shows the results for the above input parameters. If the expected cost for mixed is less than specialized a (red) dot is plotted on a graph at (p_1, p_2). If the expected cost for specialized is less than mixed a (blue) "x" is plotted on the graph at (p_1, p_2). And if the two strategies expected costs are equal then a (black) "+" is plotted on the graph at (p_1, p_2). We can see that no pattern exists in Figure 5.6. This means that the best strategy for this model depends on the parameters.

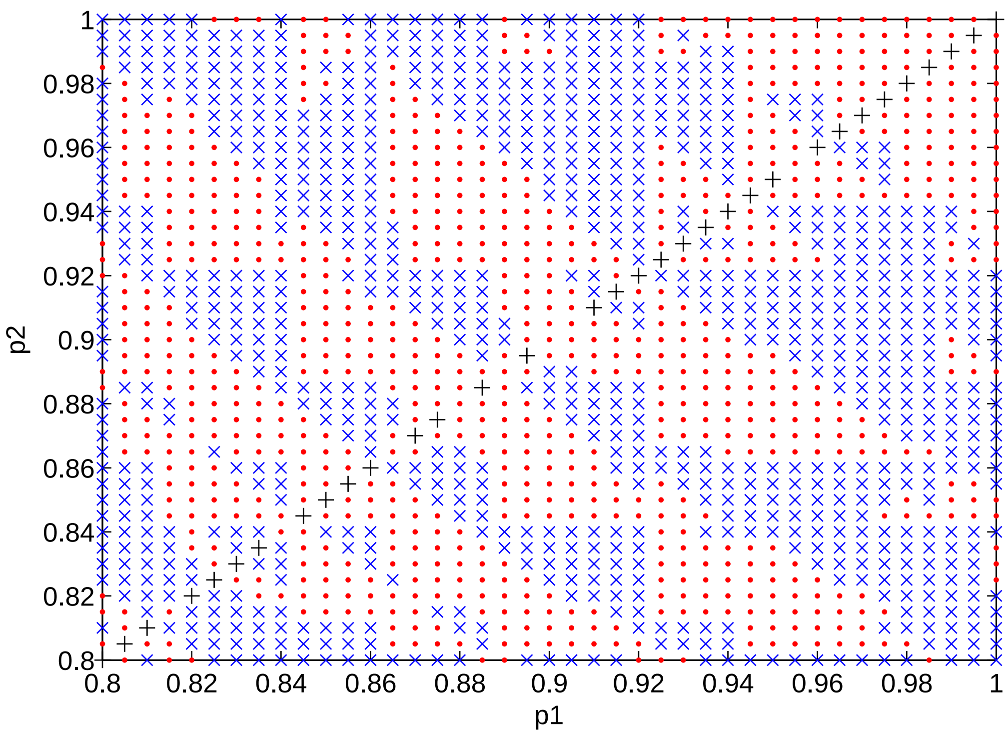

Figure 5.6: Mixed vs. Specialized - Expected Cost

Chapter 6

Conclusions and Future Research

In this thesis we develop a model formulation that investigates two different types of clinic configuration strategies to determine which is better. We compare a mixed strategy to a specialized strategy to see if the type of clinic configuration can help reduce the negative effects that patient no shows can have on a clinic's expected net revenue. Figure 1.1 shows the two different types of strategies. This investigation has shown that a health care clinic should evaluate it's system to determine which strategy is warranted. In cases where there are only two clinics/providers, overall a specialized strategy is better when the two show rates differ significantly. However, when the two show rates are close and the booking limits are similar then a mixed strategy is preferred. Similar results are also observed for multiple clinics/providers. However, in general the best strategy depends on clinic parameters.

When the two strategies were investigated further we found that there can be a significant difference in the expected net revenue between the two strategies. If a mixed strategy were to be chosen over a specialized strategy, it was found that there can be up to a 13.34% difference in expected net revenue. This means that a specialized strategy does have the ability to improve the efficiency of health care clinics under certain circumstances. This is because a specialized strategy streams patients which limits variability in no shows.

There are many ways to expand this model. First, our model only considers two types of show rates for patients, but there could be many different patient categories. Also, patient wait times could be included in the cost of overtime, because as more patients are overbooked there is the possibility the patients may spend a significant amount of time waiting in the clinic for their appointments. This lowers patient satisfaction. Another factor that could be considered is service time variability.

This paper was motivated by an analysis of appointment data from the Veterans Health Administration. However, the formulation is relevant to most clinics that experience patient no shows and the methodology used to evaluate the strategies can be used across clinics. The input parameters and outcomes would be different across different settings, but the patterns that result would be similar. Clinic configurations can make a difference to a clinic's revenues, which we used as a proxy for the clinic's multiple objectives. Therefore, it is important for clinic operating managers and researchers to include clinic configuration as one of the strategies for reducing the negative effects of no shows.

Bibliography

[1] Bean A.G. and Talaga J. Appointment breaking: causes and solutions. *Journal of health care marketing*, 12(4):14, 1992.

[2] Benjamin-Bauman J., Reiss ML, and Bailey JS. Increasing appointment keeping by reducing the call-appointment interval. *Journal of applied behavior analysis*, 17(3):295, 1984.

[3] Bigby J.A., Giblin J., Pappius E.M., and Goldman L. Appointment reminders to reduce no show rates: A stratified analysis of their cost-effectiveness. *JAMA*, 250(13):1742, 1983.

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

[5] Chakraborty S., Muthuraman K., and Lawley M. Sequential clinical scheduling with patient no shows and general service time distributions. *IIE Transactions*, 42(5):354–366, 2010.

[6] Cohen A.J., Weinstein P., and Wurster C. The Effects of Patient-Initiated Phone Confirmation Strategies on Appointment Keeping at a Hospital Dental Clinic. *Journal of public health dentistry*, 40(1):64–68, 1980.

[7] Denton B. and Gupta D. A sequential bounding approach for optimal appointment scheduling. *IIE Transactions*, 35(11):1003–1016, 2003.

[8] Gonnella J.S. and Herman M.W. Continuity of care. *JAMA*, 243(4):352, 1980.

[9] Green L.V. and Savin S. Reducing delays for medical appointments: A queueing approach. *Operations Research*, 56(6):1526–1538, 2008.

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

[11] Guse C.E., Richardson L., Carle M., and Schmidt K. The effect of exit-interview patient education on no show rates at a family practice residency clinic. *The Journal of the American Board of Family Medicine*, 16(5):399, 2003.

[12] Hassin R. and Mendel S. Scheduling arrivals to queues: A single-server model with no shows. *Management Science*, 54(3):565–572, 2008.

[13] Hixon A.L., Chapman R.W., and Nuovo J. Failure to keep clinic appointments: implications for residency education and productivity. *FAMILY MEDICINE-KANSAS CITY-*, 31:627–630, 1999.

[14] Ho C.J. and Lau H.S. Minimizing total cost in scheduling outpatient appointments. *Management Science*, 38(12):1750–1764, 1992.

[15] Kaandorp G.C. and Koole G. Optimal outpatient appointment scheduling. *Health Care Management Science*, 10(3):217–229, 2007.

[16] Kim S. and Giachetti R.E. A stochastic mathematical appointment overbooking model for healthcare providers to improve profits. *Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on*, 36(6):1211–1219, 2006.

[17] Kimbuende E., Lundy J., Ranji U. and Salganicoff A. U.S. Health Care Costs. Available at http://www.kaiseredu.org/Issue-Modules/US-Health-Care-Costs/Background-Brief.aspx, downloaded on August 19, 2010.

[18] Koren M.E., Bartel J.C., and Corliss J. Interventions to improve patient appointments in an ambulatory care facility. *The Journal of Ambulatory Care Management*, 17(3):76, 1994.

[19] LaGanga L.R. and Lawrence S.R. Clinic Overbooking to Improve Patient Access and Increase Provider Productivity. *Decision Sciences*, 38(2):251–276, 2007.

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

[21] Muthuraman K. and Lawley M. A stochastic overbooking model for outpatient clinical scheduling with no shows. *IIE Transactions*, 40(9):820–837, 2008.

[22] Patrick J., Puterman M.L., and Queyranne M. Dynamic multi-priority patient scheduling for a diagnostic resource. *Operations research*, 56(6):1507–1525, 2008.

[23] Pearce T., O'Shea J.S., and Wessen A.F. Correlations between appointment keeping and reorganization of hospital ambulatory pediatric services. *Pediatrics*, 64(1):81, 1979.

[24] Qu X., Rardin R.L., Williams J.A.S., and Willis D.R. Matching daily healthcare provider capacity to demand in advanced access scheduling systems. *European Journal of Operational Research*, 183(2):812–826, 2007.

[25] Rice JM and Lutzker JR. Reducing noncompliance to follow-up appointment keeping at a family practice center. *Journal of applied behavior analysis*, 17(3):303, 1984.

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

[27] Robinson L.W. and Chen R.R. A comparison of traditional and open-access policies for appointment scheduling. *Manufacturing & Service Operations Management*, 12(2):330–346, 2010.

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

[29] Russell L.B. Completing costs: patients' time. *Medical Care*, 47(7):S89–S93, 2009.

[30] Rust C.T., Gallups N.H., Clark W.S., Jones D.S., and Wilcox W.D. Patient appointment failures in pediatric resident continuity clinics. *Archives of Pediatrics and Adolescent Medicine*, 149(6):693, 1995.

[31] Vikander T., Parnicky K., Demers R., Frisof K., Demers P., and Chase N. New-patient no shows in an urban family practice center: analysis and intervention. *The Journal of family practice*, 22(3):263, 1986.

[32] Wang W. and Gupta D. Adaptive Appointment Systems with Patient Preferences. 2011.

[33] Yabroff R.K., Warren J.L., Knopf K., Davis W.W., and Brown M.L. Estimating Patient Time Costs Associated with Colorectal Cancer Care. *Medical Care*, 43(7):640–648, 2005.

[34] Zeng B., Turkcan A., Lin J., and Lawley M. Clinic scheduling models with overbooking for patients with heterogeneous no show probabilities. *Annals of Operations Research*, pages 1–24, 2010.