2011

Examining The Influence Of Dependent Demand Arrivals On Patient Scheduling

Husniyah B. Abdus-Salaam
North Carolina Agricultural and Technical State University

EXAMINING THE INFLUENCE OF DEPENDENT DEMAND ARRIVALS ON PATIENT SCHEDULING

by

Husniyah B. Abdus-Salaam

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of
DOCTOR OF PHILOSOPHY

Department: Industrial and Systems Engineering
Major: Industrial and Systems Engineering
Major Professor: Dr. Lauren Davis

North Carolina A&T State University
Greensboro, North Carolina
2011

ABSTRACT

Abdus-Salaam, Husniyah B. EXAMINING THE INFLUENCE OF DEPENDENT DEMAND ARRIVALS ON PATIENT SCHEDULING. **(Advisor: Lauren Davis),** North Carolina Agricultural and Technical State University.

This research examines the influence of batch appointments on patient scheduling systems. Batch appointments are characterized by multiple patients within a family desiring appointments within the same time frame. These patients are considered to be dependent amongst each other within the batch request for both arrival and no-shows. Three models are proposed to further understand the impact of these dependent demand arrivals. First, a multivariate statistical model is developed to understand the behavior of patients at public and private dental clinics. Results indicate that approximately 42% of all appointments are associated with a batch request. Also, there is a dependency among patients that are scheduled within the batch. Next, a stationary infinite-horizon Markov decision process is presented to determine the acceptance of batch appointment requests given that a finite number of open appointment slots have been reserved for same-day requests. Results indicate that the clinic should reject the request for a batch appointment when the expected number of patients in the system exceeds the number of available dentist and the probability of no-show is less than or equal to 0.10. In the final model, a finite-horizon stochastic dynamic programming model is constructed to understand the impact of the appointment demand types (i.e. individual versus batch) and overbooking on the total expected profit and the total number of patients that are overbooked. As a result, the scheduling coordinator should consider accepting batch appointments as overbooked rather than prescheduled patients.

School of Graduate Studies
North Carolina Agricultural and Technical State University

This is to certify that the Doctoral Dissertation of

Husniyah B. Abdus-Salaam

has met the dissertation requirements of
North Carolina Agricultural and Technical State University

Greensboro, North Carolina
2011

Approved by:

Dr. Lauren Davis
Major Professor

Dr. Xiuli Qu
Committee Member

Dr. Bala Ram
Committee Member

Dr. Julie Ivy
Committee Member

Dr. Paul Stanfield
Committee Member

Dr. Paul Stanfield
Department Chairperson

Dr. Sanjiv Sarin
Dean of Graduate Studies

BIOGRAPHICAL SKETCH

Husniyah B. Abdus-Salaam was born August 31, 1982 in Buffalo, New York. She received her Bachelor of Science degree in 2004 from North Carolina Agricultural and Technical State University in Electrical Engineering. She continued her education at the University of North Carolina at Charlotte, where she received her Master of Science degree in Engineering Management in 2006. She is a candidate for the Doctor of Philosophy in Industrial and Systems Engineering.

ACKNOWLEDGMENTS

Special thanks to my family and friends for their continual love and support as I have matriculated through the doctoral program. Thanks to my mother, for her words of encouragement and infinite wisdom. To my advisor, I am delighted to be her first doctoral student. Thanks for challenging and encouraging me to step beyond my comfort zone. To my peers, research group and committee members, I am truly grateful to have them in my life. They have contributed to my success in their special way. To Dr. Steven Jiang, Daniel de Oliveira Mota and Ruben Buaba, I appreciate all their help in assisting me with developing my models. I also want to thank Tanina Bradley and Adam Bowen for being my anchor on a daily basis. I could have never made it through this time without them.

In closing, this work was funded and supported by the United States Department of Education Title III Fellowship and the Alfred P. Sloan Foundation. In addition, this research was reviewed in accordance with federal regulations governing human subjects research, including those found at 45 CFR 46 (Common Rule), 45 CFR 164 (HIPAA), 21 CFR 50 & 56 (FDA), and 40 CFR 26 (EPA), where applicable.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURES **PAGE**

LIST OF TABLES

TABLES **PAGE**

LIST OF NOTATIONS

C	Capacity of the clinic
M	Maximum overbooking limit
i	Number of physicians busy due to prescheduled appointments
j	Number of patients assigned to the backlog
a	Number of prescheduled appointments assigned to the next period
b	Batch size of appointment request
n	Number of no-show/cancellations
x	Number of appointments required
y	Number of appointments available
$p_{N\|i}(n)$	Probability of no-show of prescheduled appointment given current state i
$p_B(b)$	Probability of arrival of request for batch appointment
$p_A(a)$	Probability of prescheduled appointment assigned to next period
$p(s'\|s)$	Probability of transitioning from current state, s, to future state, s'
s	Current state of clinic for two-dimensional space (i, j)
s'	Future state of clinic for two-dimensional space (i', j')
π_s	Steady-state probability for state s
$r(s,k)$	Expected immediate reward vector
k	Binary value to determine the acceptance of a batch appointment request
δ	Revenue generated from serving a patient
λ_1	Penalty cost associated with carrying a backlog

λ_2	Penalty cost associated with having unutilized appointment slots
$E[S]$	Overall expected number of patients served
$E[PS\|i]$	Expected number of prescheduled patients served
$E[BS\|j]$	Expected number of backlogged patients served
$E[b]$	Expected number of patients backlogged
U	Utilization of physicians

CHAPTER 1

Introduction

1.1 Background

In the United States, two of the most critical problems that have been encountered by outpatient clinics are the inefficiency of healthcare delivery and the inability to access timely care for patients seeking service. These problems entail high healthcare cost (approximately $2.5 trillion in 2009 or 17.6% of the nation's Gross Domestic Product) and poor healthcare quality in which the industry is forced to reduce the healthcare cost with the aid of government programs like Medicaid and Medicare. Healthcare insurance is not mandatory in the United States [1]. Based on the data from 2003, 43% of adults with chronic disease or poor health are either uninsured or underinsured in the United States [2].

Although, Medicaid and Medicare programs help with insuring individuals, approximately 15 percent of the United States population are still uninsured (roughly 43 million individuals) [3]. As a result, the uninsured patients present additional concerns for healthcare providers based on the following reasons: they often receive a lower quality of service due to the delay in obtaining necessary care and their inability to pay for better treatment options; they tend to misuse the emergency department due to a lack of a primary care physician; and they fail to obtain follow-up care due to a lack of support and resources (i.e. physicians, treatment plans/materials, etc) [3]. As stated by the Centers for Medicare and Medicaid Services, "the Medicaid Program provides medical benefits to groups of low-income people, some who may have no medical insurance or inadequate medical insurance."

Understanding the Medicaid program is essential, since this research is motivated by a government operated dental agency. With the assistance of the federal government, states must determine the eligibility of persons based on being categorically needy, medically needy or within other special groups. For example, categorically needy is represented by persons who are at or below the federal poverty level, which includes children between the ages of 6 to 19, pregnant women, caregivers and Supplemental Security Income (SSI) recipients. These individuals can also be considered medically needy; if they are above the federal poverty level, but are unable to afford medical coverage. Furthermore, the State Children's Health Insurance Program (SCHIP) provides medical care, under the Medicaid program, to children whose parents are unable to afford private insurance, but income is above the federal poverty level. In addition, these programs are developed to combat the challenges that stem from infant mortality, increases in hospitalization rates, and frequency of physician visits for persons living in poverty [4].

In their effort to redesign the healthcare system, the National Institute of Health (NIH) has identified six key performance goals. These goals consist of the following: safety, effectiveness, focus on patient-centered care, timely and efficient care, and health services that are equitable from patient to patient. [5] With these objectives, the healthcare system can sufficiently eliminate or reduce patient's behaviors in relation to calling for earlier appointments, going to the emergency department, requesting specialty consultations and prescriptions by telephone, cancelling of existing appointments, and walk-ins.

Inadequate access to healthcare providers is one of the most dissatisfying truths about the healthcare industry. Often, patients must wait an average of at least three weeks to see their physician [6]. Patient dissatisfaction is directly linked to the inadequate access to primary care physicians. Ross and Patrick [7] conducted a focus group to identify patient's attitudes towards the healthcare system. From this study, the following experiences were identified: difficulty in scheduling timely appointments, long waits during and prior to an appointment, the inability to see their primary physician, and the inability to address multiple concerns in a single visit. The main cause of these problems is the traditional scheduling system, and its inability to optimize appointment scheduling by segmenting appointment slots into prescheduled and same-day. The traditional approach requires patients to make appointments several weeks to months in advance after their visit. The push and pull between non-urgent verses urgent appointments, results in loss of timely care, and undermines healthcare quality. In addition, the long waiting time for appointments results in missed appointments and reduced efficiency of the overall daily operations. This scheduling system is also designed based on physicians' preferences, which often creates a barrier in physician-patient interactions.

Open-access, also referred to as advanced access or same-day appointment scheduling, transforms traditional scheduling systems into prescheduled and same-day appointments. This system shifts to a patient-centered model that aims to provide timely access to care and improve continuity of care, while allowing patients to see their primary physician within the same-day of the request for an appointment [6, 8]. With the newly implemented open-access scheduling system, healthcare providers have seen a significant increase in revenues as they are faced with the dilemma of how to manage capacity while meeting customers demand [9]. Due to the perishable nature of appointment time slots, revenue management will allow healthcare providers to take advantage of advanced and same-day scheduling of appointments in the most profitable manner.

O'Hare and Corlett [9] have identified several benefits associated with the implementation of open-access scheduling. The benefits experienced by the patients consisted of their ability to see their own physician and more efficient and effective visits. As a result, patient satisfaction improved significantly. The clinics noticed an increase in the physician's compensation, which leads to a higher net gain for the clinics. In addition, the clinics operated more efficiently and experienced a decreased use of urgent-care services. However, implementers of open-access scheduling have also experienced challenges that may entail one of the following: idle time of physicians when demand is low; lack of alliance between patient and physician, where the patient is more responsible for initiating and maintaining their healthcare; lack of willingness of physicians to shift control to patients for managing care; patients who are resistant to change; and/or the overuse of open-access system for patients, who tend to abuse the ability to see a physician at their own time [10]. Thus, understanding the fundamentals of revenue management is vital to examining how it is possible for open-access scheduling system implementers to generate an increase in revenues as they transition from traditional scheduling systems.

Revenue management (RM), also known as yield management, is a process that anticipates and reacts to consumer spending behavior in order to maximize revenue or profits. Pak and Piersma [11] defined RM "as the art of maximizing profit generated from a limited capacity of a product over a finite horizon by selling each product to the right customer at the right time for the right price." Weatherford et al. [12] suggest redefining the terms revenue and yield management to perishable asset revenue management (PARM). They argue that PARM will help identify the optimal tradeoff between average price paid and capacity utilization in other industries besides the traditional use in the airlines industry. In general, yield management is a comprehensive system that incorporates many of the strategies that entail reservation systems, overbooking, and segmenting demand. These strategies are applicable to service firms that have relatively fixed capacity, ability to segment markets, perishable inventory, products sold in advance, fluctuating demand, low marginal sales cost and high marginal capacity change costs [13]. RM not only increases profitability, but also allows efficiency in scheduling, which can apply to clinics in determining appointment schedules. When time slots are not filled, it costs clinics both time and money.

There are two typical tactics associated with revenue management: variation of price dynamically over time to maximize expected revenue; and overbooking sales of the asset to account for cancellations. Due to the nature of the healthcare industry, only one of the tactics is applicable in which the practice of overbooking appointments is considered a norm. Traditionally, dynamic pricing is used to vary price over time for a perishable asset. The asset owner must be able to estimate the value of the asset over time and forecast the impact of price on customer demand effectively. However, it is unethical for healthcare providers to dynamically change the price of service in order to maximize revenues.

Revenue management has encountered a certain myopia inside the field, where practitioners and researchers view RM solely in airline-specific terms. This has restricted both research and implementation efforts in other industries. Also, RM is viewed negatively due to airline pricing with consumers, where fares are complex; therefore, managers are reluctant to try RM practices. “Applying RM does not involve radically changing the structure of pricing and sales practices; rather, it is a matter of making more intelligent decisions [14]. ”

Over the past decade, there has been an increase in revenue management techniques in the service industry. This movement can be attributed to new approaches in how decisions are made. These methods should be technologically sophisticated, detailed, and intensely operational. Due to the advances in economics, statistics, and operations research, it is possible to model demand and economic conditions. Researchers are also able to quantify the uncertainties faced by decision makers by estimating and forecasting market response. This allows researchers to compute optimal solutions to complex decision problems. Information technology has provided the capability to: automate transactions, capture and store vast amounts of data, quickly execute complex algorithms, and implement and manage highly detailed demand-management decisions. These advances have also lead to problems involving the possibility of managing demand on a scale and complexity that would be unthinkable through manual means and the possibility of improving the quality of demand-management decisions [14].

1.2 Research Scope

This research aims to provide effective, patient-centered care, in a timely and efficient manner through a well-developed open-access scheduling system. Influenced by the benefits of open-access implementation, this research also aims to model how this scheduling system allows clinics to increase their revenues. Current research in applying revenue management to primary-care clinics has modeled the effect of patient choices in deciding on whether to accept or reject a same-day appointment [15]. The research presented in this paper examines the impact of dependent arrivals on an open-access scheduling system while maximizing revenue. To the author's knowledge, quantitative models, developed to understand open-access scheduling systems, only explore single independent patient arrivals within predominately single provider scheduling models. However, there are a few journal papers that study the impact of open-access scheduling on multiple provider models.

The objective of this paper is to introduce dependent demand arrivals in relation to multiple providers modeling. Theoretically, the patients are considered dependent due to the fact that knowledge that a patient within a group (batch) will not meet their scheduled appointment or cancels, affects the probability that the other patients within the same group (batch) will also not meet their scheduled appointment or cancel. Thus, the arrival of patients is dependent among each other within the batch. A batch accounts for two or more individuals who are interested in scheduling appointments with a set of providers, given that each person within the batch will be served. This research also aims to understand the challenges faced with both public and private pediatric dental clinics in scheduling batch appointment requests.

When faced with limited resources, it is essential for policy makers to use effective methods in planning, prioritization, and decision making [16]. Therefore, three models are presented in this paper to assist scheduling coordinators in examining the effects of batch appointments on their scheduling paradigm. First, this work identifies the prevalence of batch appointment requests at public and private pediatric dental clinics. In the case study, data analysis and multivariate statistical analysis are used to answer the following questions:

1. How does the prevalence of batch appointments differ based on clinic type?
2. Are patients scheduled within a group dependent amongst each other in terms of their arrival and no-show rate?
3. Is there a relationship among appointment demand type (individual versus batched patients), patient behavior (break or meet scheduled appointment), and reason for the appointment?
4. Which variables predict the behavior of the patient?

Next, a stationary, discrete time, infinite-horizon Markov decision process (MDP) model is developed to equip the scheduling coordinator with better information to be used in identifying the optimal policy in the acceptance and rejection of batch appointment requests. With this model, the following questions are addressed:

1. How is the open-access scheduling paradigm, in terms of the percentage of the appointment slots that are allocated to same-day request, affected by the batch appointment requests?
2. How is the optimal scheduling policy affected by varying degrees of the percentage of prescheduled patients, patient behavior, and appointment request size?
3. How does overbooking affect the performance of the clinic?

The behavior of the system is quantified under several performance measures including: the total expected number of patients that are served, the utilization of physicians, and the expected number of patients assigned to the backlog (i.e. overbooked). The backlog represents the queue of patients that are waiting in the system to be seen by the first available physician. Based on the results of these questions, a complete analysis of the scheduling system is generated. Thus, allowing us to compare the predetermined performance measures for various no-show rates and prescheduled appointments ratios.

In the final model, a finite-horizon stochastic dynamic programming model is constructed to study the impact of appointment demand types (individual versus batched patients) on the clinic's profitability and the physicians' productivity. More importantly, this research determines if batch appointments negatively impact the clinic's performance, in terms of the number of patients accepted and the utilization of physicians. Also, the optimal scheduling policy is identified, which is based on the acceptance of individual patients only, batch appointments only, or a hybrid of both demand types. With each of these models presented hereafter, healthcare providers gain insight into not only the prevalence of batch appointments, but also the influence of these appointment requests on patient scheduling systems.

1.3 Dissertation Overview

The paper is composed of five remaining chapters. Chapter 2 provides an in-depth literature review of revenue management and appointment scheduling systems. Chapter 3 presents a case study on the prevalence of batch appointments for public and private pediatric dental clinics. Chapter 4 examines the acceptance of batch appointments in a discrete-time, discrete-space, stationary infinite-horizon Markov decision process model. Chapter 5 explores the effects of scheduling independent versus dependent patients under a finite-horizon stochastic dynamic programming model. Chapter 6 summarizes and concludes the research presented here. In addition, possible extensions of this work are presented.

CHAPTER 2

Literature Review

2.1 Introduction

In 2009, the cost of healthcare was approximately $2.5 trillion or 17.6% of the nation's Gross Domestic Product [1]. However, the growth of spending declined by 4.0 percent from the previous year. Figure 1 illustrates the total spending trend from 1960 to 2009 in the US, as well as the spending trends for both the private and public sectors and out-of pocket. The significant rise in healthcare costs can be attributed to such factors as increased technological costs, an aging population with health problems, defensive medicine, excess capacity, and an increased number of well-trained specialists demanding higher wages [17]. With surging healthcare cost, the United States is currently identifying methods to reform the existing healthcare system.

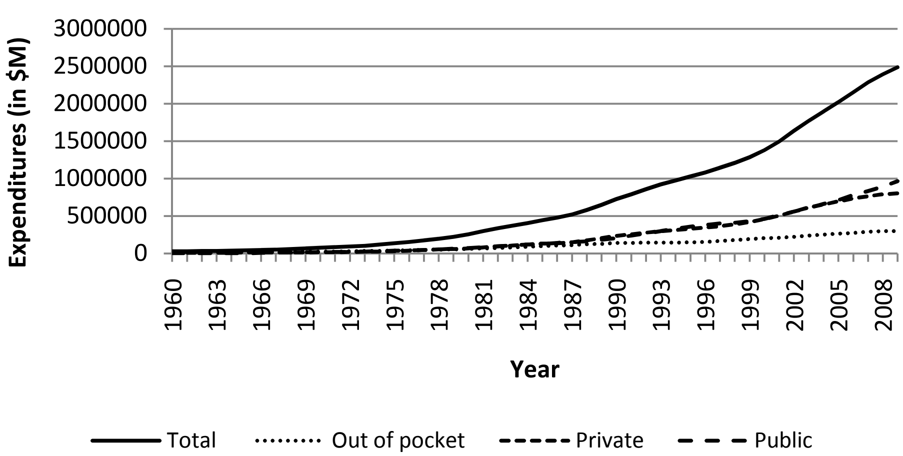


Figure 1. National healthcare expenditures in millions of dollars

With respect to personal healthcare, the Centers for Medicare and Medicaid (CMMS) have categorized the type of medical service and product into the following: hospital, professional services, nursing home and home health, and retail outlet sales of medical products. Professional services entail physician and clinical service, other professional services (i.e. therapists and chiropractors), dental services, other personal healthcare in nontraditional settings (i.e. school and community centers); whereas medical products are composed of prescription drugs, durable and non-durable medical equipment. Figure 2 breakdowns the healthcare expenditures by the type of service or product for 2009. Hospitals and physicians and clinical services account for the majority of the healthcare cost at approximately 65%. We focus our attention to the dental spending, which only makes up about 5% but has a lower percent change from the previous year than physicians and clinical services. This is evident in Figure 3 which represents the percent change for the overall expenditures, physician and clinical services, and dental services from 1970 to 2009.

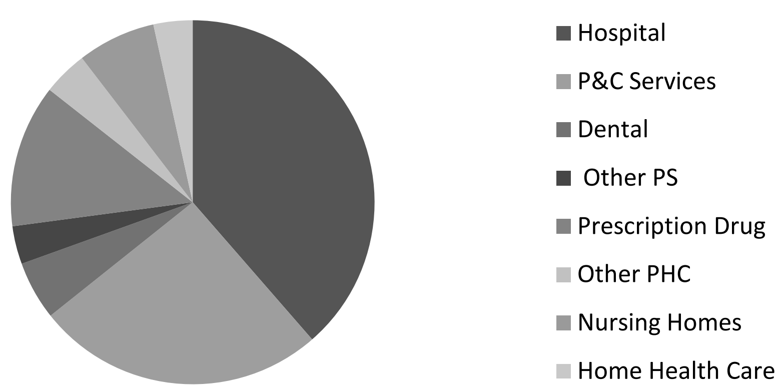


Figure 2. Expenditures for personal healthcare

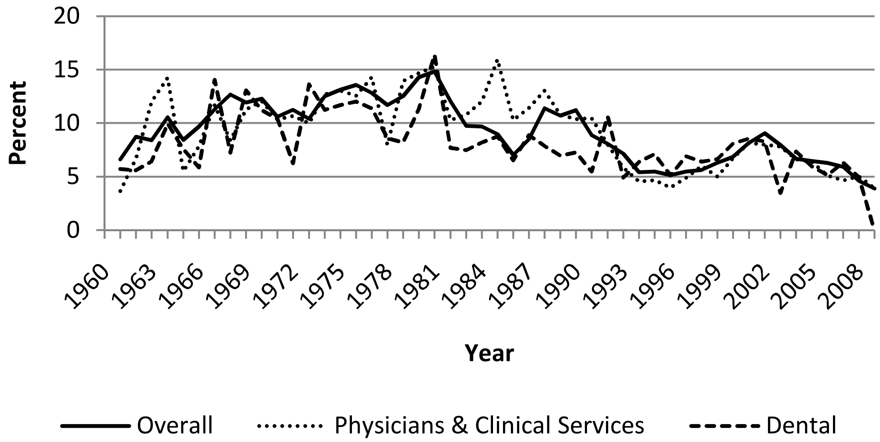


Figure 3. Expenditures change from previous year for professional services

Healthcare administrators must determine ways to combat this rising healthcare cost. As with any professional healthcare service, the dental service must determine methods to minimize cost/maximize revenue and maximize utilization of personnel, equipment and other resources. The most efficient method must embody a technique that can balance supply and demand for service. Therefore, this research examines the use of revenue management and open-access scheduling techniques.

The objectives of this chapter is to (i) review literature on revenue management, (ii) review literature on patient scheduling and open-access scheduling, (iii) identify how open-access scheduling is associated with revenue management problems/techniques, and (iv) determine how healthcare is unique to traditional RM industries.

The remainder of this chapter is composed of five remaining sections. First, the boundaries of the study are discussed. In Section 3, an introduction to revenue management is presented based on its characteristics, techniques/models, applications and research opportunities. In addition, this research explores the use of revenue management within the healthcare industry and its uniqueness. Then, this work examines the general concept of appointment scheduling and more specifically overbooking models and open-access scheduling systems in Section 4. Finally, the gaps in research are identified, along with potential research directions and challenges.

2.2 Study Boundaries

A comprehensive discussion on the history of revenue management is given by McGill and Van Ryzin [18], in which the authors propose several possible directions where RM can be utilized. Chiang et al. [19] published an overview of the literature related to revenue management. In this journal paper, the authors address the issues and potential research directions of RM. They present insight of the progress of RM since 1999, with the understanding of customers' value functions and behavior essential in designing service packages for different market segments such as walk-ins, no-shows, cancellations, appointment scheduling patterns. Thus, the goal is to identify how RM has been used in nontraditional industries and understand how RM techniques can be applied to an open-access scheduling system. In order to accomplish this goal, a literature search is conducted on both revenue management and open-access scheduling.

Throughout the literature search process, several questions will be raised: What is revenue management? What are the characteristics found in traditional verses nontraditional industries? How is the healthcare industry different from traditional industries? What is open-access scheduling? How does open-access scheduling differ from other scheduling models? How can RM be applied to open-access scheduling? The proposed questions are answered in the following sections.

2.2.1 Definitions of key concepts

The broad term "revenue management" refers to the wide range of techniques, decisions, methods, process, and technologies involved in demand. The demand-management decisions consist of three basic categories: structural decisions, price decisions, and quantity decisions [20]. As defined by Chopra and Meindl [21], revenue management is the use of pricing to increase the profit generated from a limited supply of assets in the form of capacity and inventory.

Moreover, open-access scheduling is a patient-centered system that aims to provide timely access to care and improve continuity of care, while allowing patients to see their primary physician within the same-day of the request for an appointment [6, 8].

2.2.2 Search process

An extensive literature review is conducted on revenue management and open-access scheduling. Several databases are used in the search process including Google Scholar, Web of Science, Compendex, and Knovel. For the revenue management literary search, several keywords are used which consist of the following: "revenue management", "yield management", "revenue management and healthcare", "revenue management and scheduling", "revenue management and patient scheduling", and "revenue management and nontraditional industries." With the open-access scheduling search, "advanced-access scheduling" and "same-day appointments" are also used as keywords.

2.2.3 Characteristics of articles

For an overview of the trends and characteristics of the revenue management literature, McGill and Van Ryzin [18] and Chiang et al. [19] are suggested. From the open-access scheduling literature search, the papers are divided into four categories. Table 1 summarizes these categories in terms of the conceptual or framework papers, case studies on the implementation of open-access scheduling in a variety of clinics, outcomes of open-access scheduling, and quantitative models.

Table 1. Open-access scheduling literature listed by categories

Conceptual

Implementation: Murray and Tantau [22], Schneck [23], Murray [24]
Potential benefits and challenges: Herriott [25], Pinto [10], Gupta and Denton [26]
Panel size: Savin [27], Green et al. [28], Murray et al. [29]
Other: Kilo and Endsley [30], Murray and Berwick [31], Randolph [32], Gupta et al. [8], Miller [33]

Implementation

Primary care: Murray et al. [34], Bundy et al. [35], Knight et al. [36]
Military: Meyers [37], Armstrong et al. [38]
Academics: Kennedy and Hsu [39], Steinbauer et al. [40]
Other: Gill[41], Belardi et al. [42], Newman et al. [43]

Outcomes

Benefits: O'hare and Corlett [9],
Challenges/barriers: Solberg et al. [44], Terry [45], Mehrotra et al. [46], Ahluwalia and Offredy [47]
Statistical: Pickin et al. [48], Parente et al. [49], Bennett and Baxley [50]
Other: Lewandowski et al. [51], O'Connor et al. [52], Fine and Busselen [53], Sperl-Hillen et al. [54], Randolph et al. [55]

Quantitative

Simulation: DeLaurentis et al.[56], Giachetti et al. [57],
Mathematical: Kopach et al.[58] , Qu et al. [6], Gupta and Wang [15], Muthuraman and Lawley [59], Qu and Shi [60]

For the conceptual papers, the respective authors introduce the framework and objectives of open-access scheduling; whereas, other papers are developed to discuss the process of implementing the scheduling system in healthcare clinics. The outcomes papers describe what happens as a result of implementing the scheduling system in a variety of healthcare clinics. Some papers provide statistical analysis to examine the benefits of open-access scheduling. Finally, the quantitative papers represent a variety of mathematical models that attempt to provide logical reasoning to support the concept of open-access scheduling. It is worth noting that several papers touch on one or more of the four categories; however, the paper is placed in the category that the author(s) focuses most of their attention. Figure 4 illustrates the distribution of these articles across the four categories.

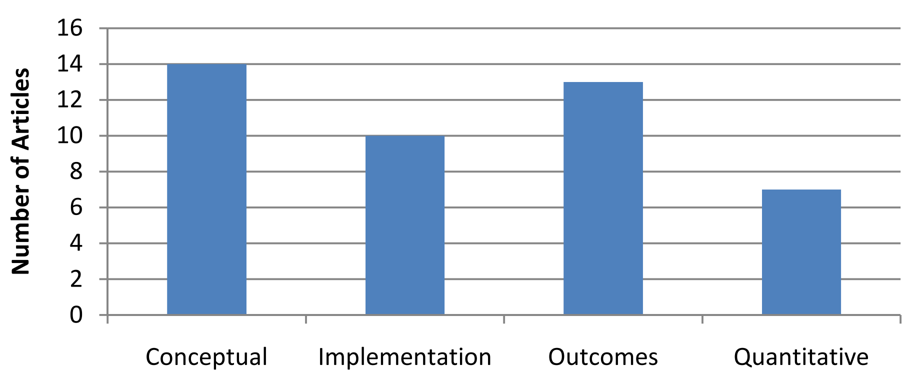


Figure 4. Literature decomposition for each category

2.3 Revenue Management

The concept of revenue management evolved due to the deregulation of United States airlines industry in the 1970s. "It was developed as an outgrowth of the need to manage capacity sold at discounted fares, which were targeted to leisure travelers, while simultaneously minimizing the dilution of revenue from business travelers willing and able to pay full fares [61]." This differential pricing strategy can be accomplished in terms of customer segment, time of use, and product or capacity availability [21]. The business environment in which RM is primarily used can be characterized as follows: 1- the existence of price variations for each market segment; 2- highly perishable inventory or capacity that can lead to wastage; 3- the presence of seasonal demand peaks (or any other form); and 4- the possibility of inventory or capacity to be sold in bulk and/or instantaneously for cash [21]. Based on these four characteristics, healthcare is applicable for revenue management since there is a high level of perishability with resources (i.e. appointment slots) and the presence of seasonal demand. Seasonal demand occurs when the number of patients seeking treatment fluctuates in the time of day, the day of the week, and the time of year. It is worth noting that perishable inventory and capacity cannot be utilized after a certain period of time. These characterizations provide the foundation for a successful revenue management model in which the right resource is sold to the right customer at the right price and time.

Under RM, demand-management decisions must be made. These decisions consist of three basic categories: structural decisions, price decisions, and quantity decisions. Structural decisions determine which selling format to use, segmentation or differentiation mechanisms, bundle of products, etc, whereas price decisions examine how to set posted prices, individual-offer prices, and reserve prices; and how to markdown over product's lifetime. As for quantity decisions, the decisions are based on whether to accept or reject an offer to buy; how to allocate output or capacity to different segments (i.e. walk-ins, prescheduled, open-access appointments); when to withhold a product from the market and sale at later points, etc. In order to determine which decision is more relevant depends on the context of the industry and the time in which the decision must be made [14]. Although RM addresses all three categories; structural decisions are considered to be strategic decisions that are taken infrequently. Thus, a greater emphasis is placed on the operational decisions using quantity-based RM (capacity-allocation decisions) and/or price-based RM (prices used to manage demand). In the research presented in this paper, we focus our attention to the quantity-based revenue management decisions as it relates to patient scheduling.

Additionally, revenue management can consist of pricing, auctions, capacity and inventory control, overbooking, and forecasting models. With pricing models, one must determine the price for various customer groups and how to vary prices over time to maximize revenues or profits; whereas auctions are used to address methods for dynamically adjusting prices. Capacity and inventory control determines how to allocate capacity of a resource or a bundle of different resources to different classes of demand, so that the expected revenue or profit is maximized. The overbooking model is used to increase the total volume of sales by selling reservations above capacity to compensate for cancellations and no-shows. Finally, forecasting is essential in the quality of RM decisions based on pricing, capacity control, and/or overbooking. Forecasting can be done using full- and semi-aggregated and fully disaggregated models [19]. In regards to the healthcare industry, a mixture of these models can be used from capacity/ inventory control, overbooking and forecasting. It is questionable if pricing and auctions can be applied due to the complexity of the healthcare's billing and reimbursement practices. Figure 5 illustrates open-access scheduling in relation to the revenue management models. The highlighted portions of the diagram represent those models that are applicable in the healthcare industry.

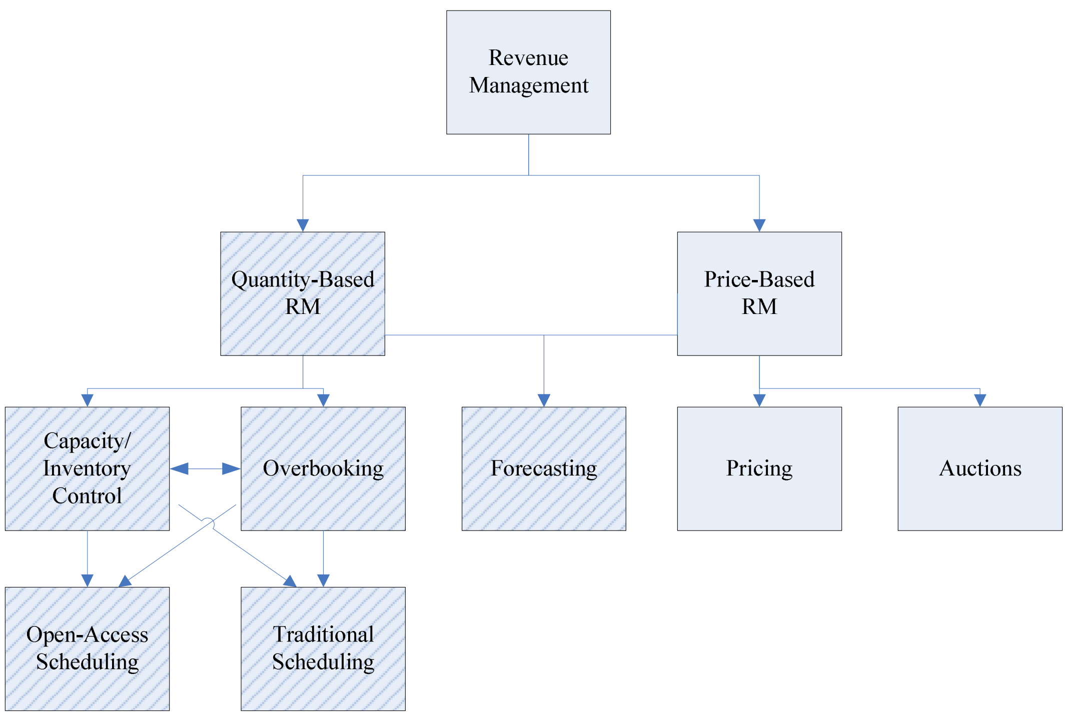


Figure 5. Revenue management problems with respect to scheduling

2.3.1 Applications of revenue management in the service industry

Not only has the airline industry utilized the RM concept, but the hotel [62-63], rental car [64], air cargo [65], professional service firms [66], nonprofit businesses [67], project management [68], restaurants [69-70], retail and manufacturing [71-74] industries as well. These industries all have similarities in that they face high fixed costs/low variable cost, spoilage and temporary demand imbalances [13]. Chiang et al. [19] suggest that advancements in information technology (IT) have led to more sophisticated RM capabilities. Some of the nontraditional industries identified include the following:

- Hospitality- restaurants, hospitals and healthcare, attractions, cruise lines, casinos, saunas, resort, golf, sports events, conferences, etc
- Transportation- boat, railways, cargo and freight
- Subscription services- IT services and internet services, cellular networks, television services
- Other industries- retailing, manufacturing, broadcasting and media, natural gas, project management, apartment renting, sales management, inclusive holiday, nonprofit sector. [19]

Chiang et al. [19] pose several questions in regards to future research: how to apply RM in nontraditional industries; how to use new methodologies such as auctions, e-commerce and internet marketing to improve the performance of RM; how to make RM decisions more effectively under competitive and collaborative environment; and how to make forecasts more accurately. The authors also provide insight into how customers' value functions and behavior are essential in designing service packages for different market segments such as walk-ins, no-shows, cancellations, appointment scheduling patterns. So what can the healthcare industry learn from other industries? Table 2 provides the lessons learned from traditional industries that have applied the revenue management concept based on the information provided by Chiang et al. [19].

Table 2. Best practices from traditional industries

Industry	Lessons Learned
Airlines	Utilize both overbooking and segmentation
Hotel	Use overbooking policy to compensate for cancellations and no-shows
Resorts	Segment customers based on scheduling types
Rental Cars	Decide whether to accept or reject booking requests based on length-of-rent controls
Cargo & Freight, IT & internet services	Utilize capacity planning techniques

2.3.2 Healthcare uniqueness

Bell stated in 1998, his general belief that "RM concepts will soon be applied to almost everything that is sold." Each new industry introduces a new set of challenges and a new perspective to RM. By shifting the focus from relying on capacity/inventory controls, Karaesmen and Nakshin [2] suggest hospitals consider pricing optimization to attain financial goals. Government and private insurance covers the expense of most hospital services. Roughly 33% of hospital's revenue comes from Medicare, about 33% from commercial insurance and 17% from Medicaid; the remainder comes from out-of-pocket and charitable care.

Hospital pricing and billing practices are highly complex. As a result, the healthcare system is considered a customized service with the use of charge master pricing and contract prices to aid in billing. For example, a patient is presented a charge master at the time of their billing which provides the net price for each item of service and/or products that have been used during their healthcare service. This price does not account for the actual amount that will be covered by the patient's insurance coverage and/or the amount the patient will be responsible for (if any) [2].

In order for revenue management to be effective in the healthcare industry, several underlying problem characteristics must be considered that are unique to this industry. In traditional industries, it is common to assume the capacity is fixed. However, a medical clinic can easily increase capacity through the use of overtime. Within the healthcare industry, segmentation is not practiced in a manner that allows for revenues to be maximized. Also, healthcare providers are often faced with managing a backlog of appointments. Finally, the healthcare industry has very different cost structures than those found in hotel and airline industry [75].

Unlike other industries, healthcare is faced with managing reimbursements and revenue cycles to be financially viable. Revenue cycle management involves the payment for a product or service that is not made in advance or immediately at the time of sales/service. Healthcare uniqueness from other industries entails customers who cannot opt out for not using a medical service, in many cases; neither can hospitals reject selling their service to the patient. RM practitioners must consider expected reimbursements from the payers and patients' ability to pay instead of their willingness to pay. Also, costs differ per customer and there is uncertainty present in identifying how much money will be reimbursed by each patient's insurance provider. The ultimate goal of pricing and revenue optimization in healthcare is to attain service goals while effectively managing finances.

Current pricing and revenue optimization (PRO) models assume variables are known or fixed. The authors suggest segmentation based on a need to determine good estimates of reimbursements and costs for capacity management and pricing optimization rather than based on patients financial situation. PRO characteristics visible in healthcare include the following: capacity related problems, high fixed cost, wealth of historical data (potential to obtain data), high transaction volume, and the ability to do price segmentation [2]. For example, Qi and Yan [76] examine the use of RM for capacity control in a community hospital to determine the optimal reserve capacity for advanced and common wards.

Within the healthcare system, the industry's infrastructure is not designed to increase the price for healthcare in order to deter customers to not use the service. Thus, the healthcare industry is often forced to cut supply of service due to the lack of resources. Several approaches are identified to handle the many varieties of demand in the healthcare industry. For example, some pressures may be best met, not by curtailing demand, but by coping with it and meeting it in a radically different way. In some instances, healthcare demand has forced some physicians' practices to incorporate a helpline to be accessed by patients to deter from unnecessary in office visits. This practice of deterring demand has reduced the increase in demand for health facilities by 40% [20].

Capacity management attempts to determine how to allocate scarce resources among different patient groups by matching demand with supply [77]. Additional service-related goals consist of increasing the number of appointments that can be scheduled, or decreasing the waiting times and delays in the healthcare system, which may neglect the hospital's financial goals [2]. Smith-Daniels et al. [17] discuss previous research and future research in the area of capacity management in healthcare services. Previous research has disclosed that trends toward growth and integration in healthcare organizations have been invalid as earlier research was performed during a time when health care essentially was characterized as a cottage industry. Within the cottage industry, patients often use a single physician or a small group of physicians throughout their entire lifetime, from birth to death, to serve their primary care needs. Given these changes in the healthcare environment, it seems appropriate to assess previous research on healthcare capacity planning and management and to determine its relevance to this changing industry.

In order for capacity management to be successful, one must determine the most effective and efficient approach in work force management and scheduling. Work-force capacity is a function of the number of personnel hours available per unit of time and the composition of the work force in terms of the mix of employee skills. Most health care organizations determine the number of full- and part-time employees of various skill levels through the annual budgeting process. As shown in most work force models, work-force acquisition decisions must consider such factors as (1) the stochastic nature of demand, (2) the difficulties in measuring the productivity of health care providers, (3) the flexibility facilitated by the substitution of different employee types, (4) the use of part-time employees for lowering operating costs and improving schedule flexibility, and (5) the use of overtime and temporary employees to provide additional work-force capacity [17].

Gupta and Wang [15] suggest that a clinic must manage patients' access to physicians' slots in order to balance the needs of those who book in advance and those who require a same-day appointment. The authors also insist that one must decide which appointment requests to accept in order to maximize revenue. Their research identifies the disadvantages in scheduling too few appointments with an increase in patients' wait time, the patient and primary care provider (PCP) mismatch, and the possibility of unutilized clinic appointment slots.

Based on the literature review, revenue management is composed of both capacity planning and demand management strategies. The primary objective is to identify the strategy that best balances the demand with the available supply in a manner that maximizes revenue. Thus, this research examines how patient scheduling systems are utilized to handle this problem.

2.4 Healthcare Appointment Scheduling

Appointment scheduling has been highly and extensively studied since the early 1950s, beginning with the research presented by Bailey [78]. In general, appointment systems are designed to minimize waiting times for patients while maximizing the utilization of physicians and other resources [79-81]. These systems can be divided into two categories: static, where decisions are made prior to the beginning of a clinic session, and dynamic, where decisions are continuously updated based on the current state of the schedule [82]. Cayirli and Veral [82] provide an extensive literature review that describe the fundamental factors associated with appointment systems that entails the following: number of services available, number of physicians, number of appointments per clinic session, the arrival process of patients in relation to punctuality, no-shows, walk-ins and presence of companions, service times, lateness and interruption level of doctors, and queue discipline. The authors also present the measures of performance in regards to cost-based, time-based, congestion, fairness, and other appropriate measures. England and Roberts [83] also suggest that performance measures should be based on less quantifiable parameters like improving community health.

With the designing of appointment systems, decisions must be made on the appointment rules (i.e. block-size, begin-block, and appointment interval), patient classification (used to determine booking sequence of patients and/or adjustment of appointment intervals to meet specific patient characteristics type), and adjustments of no-shows, walk-ins, urgent patients, emergencies, and second consultations into the system. Furthermore, appointment systems research is directed to analytical models using queueing theory and mathematical programming methods, simulation-based models, and case-studies. The authors also suggest that there is a significant gap between the theory of appointment systems and application of these systems in the actual-world [82].

Gupta and Denton [26] state the importance of appointment scheduling with respect to efficiency and timely access to health services. As with primary care, appointment systems must be designed to "find a suitable match among the available time slots of providers in the clinic, provider prescribed restrictions on how available slots may be filled and patients' preferences for day/time of week as well as for a particular service provider." In addition, Kaandorp and Koole [84] further emphasize that the scheduling objective must consider the trade-off between both the physician and the patient preferences, where the physician prefers to be more productive (less idle time) and patients tend to want shorter waiting time. By using patient classification in the design of appointment systems, Cayirli et al. [82] aim to improve patients' waiting time, physicians' idle time and overtime in the absence of making trade-offs between the patient and provider. In the remaining sections, we examine the use of overbooking as appointment scheduling models has transitioned from traditional to open-access scheduling systems.

2.4.1 Overbooking

As noted by McGill and Van Ryzin [18], overbooking is the oldest and most studied revenue management strategy within the airline industry as a response to controlling the probability of denied boardings. Overbooking is also the most utilized approach of revenue management for patient scheduling in traditional appointment models. Giachetti [85] describe overbooking as a population-based policy in which patients are overbooked for any given day to help reduce the rate of no-show and appointment delay. Overbooking is more suitable for situations where customers are able to cancel orders and the value of the asset drops significantly after a deadline. The basic trade-off is to consider the wasted capacity (inventory) due to excessive cancellations or having a shortage of capacity as a result of too few cancellations in which case an expensive alternative needs to be arranged. Cost of wasted capacity is the margin that would have been generated if the capacity had been used for production or service, whereas the cost of capacity shortage is the loss per unit that results from having to go to a backup source.

Thus, the goal is to maximize profits by minimizing the cost of wasted capacity and the cost of capacity shortage. This may potentially result in the optimal level of overbooking increasing as the margin per unit increases and the level of overbooking decreasing as the cost of replacement capacity increases [21]. In other words, the objective is to determine the optimal booking limit for each time period that maximizes expected revenues, as one considers the probability of cancellations and penalties for exceeding capacity [18].

Kim and Giachetti [75] investigate how overbooking can help healthcare providers over time and to enable more efficient use of limited resources while maximizing profits. Overbooking also allows healthcare providers to balance the costs of too few patients showing up with the costs of too many patients showing up. The benefits of overbooking to patients include reduced waiting times and increased continuity of care. The authors develop a stochastic mathematical overbooking model (SMOM) to determine the optimal number of patient appointments to accept to maximize expected total profits for diverse healthcare environments. SMOM considers the probability distribution of no-shows and walk-ins to obtain an optimal solution for the number of patient appointments to be scheduled.

LaGanga and Lawrence [86] discuss how to utilize overbooking models to schedule patients for clinics efficiently, while also managing no shows. The problem is that patient no-shows are significant in many health care settings, where no-show rates can vary from as little as 3% to as much as 80%. No-shows reduce provider productivity, increase health care costs, and limit the ability of a clinic to serve its client population by reducing its effective capacity. The paper provides a source for managers to understand appointment overbooking strategies. First, the paper shows 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, the authors differentiate clinic overbooking from overbooking for revenue management in transportation services.

Second, it shows 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 the results to the specific characteristics of a clinic. Third, the authors use simulation experiments, regression analysis, and sensitivity experiments to 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. This, in turn, translates into reduced clinic costs and improved patient satisfaction and outcomes. Fourth, the paper provides managerial insights into the practical use of appointment overbooking in actual clinics and demonstrates its application in a large publicly funded mental health clinic. It also identifies situations in which overbooking is most likely to be beneficial and, conversely, in which it is likely to be counterproductive.

Appointment scheduling and no-shows management have been used in other industries such as medical practice, healthcare administration, operations management, marketing, and transportation planning. LaGanga and Lawrence [86] state that little work has been contributed on the use of overbooking to mitigate the negative impact of no-shows in appointment-oriented services such as clinical healthcare. In contrast, transportation revenue management has been extensively examined, where overbooking has been studied in terms of capacity utilization and profitability using perishable asset revenue management. However, appointment overbooking is very different from transportation services overbooking, since appointment no-shows are spread over time, while transportation no-shows all occur at a single point in time. The difference in problem structure requires quite different solution approaches to the problem of no-shows. Many academics and healthcare managers have attempted to understand the behavior of patients that fail to meet their scheduled appointment. The goal is to find a relationship between age, gender, number of previous appointments, and the lead times given for appointments to patients. This relationship can be used to help construct a probabilistic model to find the root causes of no-shows, such that they are eliminated or reduced [86].

2.4.2 Open-access scheduling

Patient scheduling has encountered several major transitions from traditional models, to carve-out models, and now open-access scheduling models. The traditional model is more relevant to the days of prescheduling and "take-a-number-and-wait" systems, where patients have experienced long waiting times to see their physician and long waiting times to schedule an appointment. The traditional model divides appointments into two categories: urgent (same-day) and non-urgent care. The traditional system entails a pattern of double-booking appointments, high no-show rates, and patients requiring multiple appointments. From the traditional model derived the concept of the carve-out model. The carve-out model reserves urgent care time in advance, which often prevents patients from seeing their own physician. This presents a major problem due to the inability to set precedence to the continuity of care and possibly the need for a second appointment with the patient's primary physician. The carve-out system tends to push non-urgent care to a future appointment date that enables dysfunctional habits in matching supply and demand [22].

Introduced by Kaiser Permanente in northern California, the open-access scheduling model aims to rebuild their system by creating an access system focused on the key healthcare product: doctor-patient relationship with respect to both the continuity of care and capacity. Kaiser Permanente experienced an average wait of 55 days for an appointment in which only 47% of the patients were able to see their own physician. This inefficient and costly system was a result of a high rate of missed appointments, and the lost income and lost opportunity of patient visits. Furthermore, the longer the delay of care; the greater the threat to quality of care [22]. Figure 6 indentifies the percentage of appointments that are prescheduled as scheduling systems transition from a traditional to carve-out to advanced-access models; whereas, Figure 7 characterizes and exposes the associated risks of each model as scheduling models has transitioned from traditional to open-access.

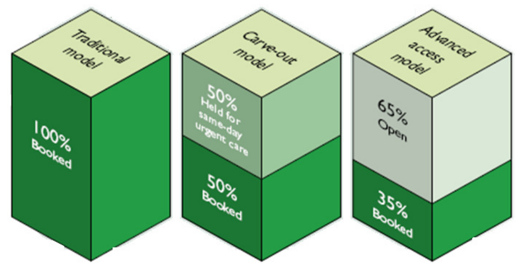


Figure 6. Scheduling models from Murray and Tantau [22]

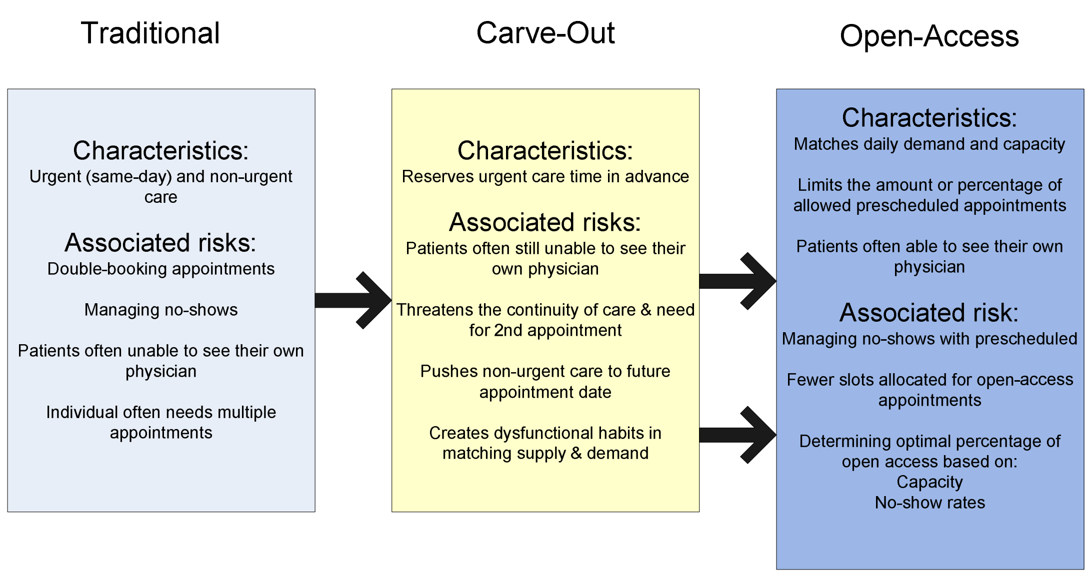


Figure 7. Evolution of scheduling models

Open-access scheduling systems allow patients to be seen at their leisure, in turn, improving healthcare delivery quality while reducing healthcare cost. How to manage capacity and meeting daily patient demand is the fundamental characteristic of this novel scheduling model. Therefore, advanced access scheduling limits the amount of prescheduled appointments for a specific timeframe. The associated risk encompasses managing no-shows with prescheduled appointment slots and undermining the capability of open-access appointments with too few appointments available. Several studies have examined the transition from traditional scheduling to open-access systems that aim to improve efficiency of primary care clinics, reduce no-show rates, manage walk-ins, reduce waiting times for scheduling appointments, and restructure types of appointments and length of appointments (refer to Table 1). Johnson et al. [87] suggest that open-access scheduling, along with patient education, patient reminders and patient sanctions, can reduce the rate at which patients fail to meet their scheduled appointments.

Several principles have been identified to assist management with the tools vital to implementing changes in their scheduling systems. The ten principles of open-access consist of:

1. Balance appointment supply with patient demand;
2. Work down the backlog;
3. Reduce appointment types;
4. Plan for contingencies;
5. Reduce future patient demand;
6. Manage the bottlenecks;

7. Synchronize patient, provider and information;

8. Predict and anticipate patient needs at the time of appointment;

9. Optimize rooms and equipment; and

10. Use continuous-flow strategies. [35]

Even with these tools, there is no guarantee that the transition from traditional to open-access scheduling systems will be challenge free. Most healthcare clinics report that managerial time is needed on a permanent basis to sustain advanced access scheduling. Several clinics also encounter trouble working down the backlog. With large organizations, problems occurred when the open-access concepts are introduced and initiated by management rather than physicians. The benefits are more visible by management (with decrease in appointments delay) than for physicians (less stressful days) which make it difficult to motivate physicians to adopt the system. The transition to open-access scheduling is easier when implemented in smaller private clinics. In addition, the lack of a contingency plan presents another challenge, if the system is unable to adapt or respond when confronted with abrupt and unexpected changes in supply and demand. Furthermore, one should focus same-day scheduling on the process and principles rather than on a specific product or solution; in order, to achieve a successful appointment system [35].

In practice, the concept of open-access scheduling has been used to develop a prediction grid that would predict actual patient arrivals based on a previous model prediction grid formulated by Kaiser Permanente. The use of open-access principles can be seen in the theory that the demand for same day appointments can be predicted and this demand prediction can help determine an actual demand of patient appointments by day of the week and by month of the year. The objective is to use open-access appointment scheduling to help improve patient's access to healthcare providers in a suitable time frame. The use of open-access techniques requires the help of management to aggressively predict patient arrivals and staffing schedules. Forjuoh et al. [88] discuss the challenge in creating a prediction grid and explain how most researchers use the Kaiser model as a backdrop because it is the only model available. Using historical data, appointment schedules for the year were generated and compared to those using the Kaiser method. The results of the experiment show that appointments scheduled by the day of the week and Kaiser method are similar, but the two approaches differed on the summer and winter month's schedules. The authors concluded that the results from the Kaiser model may be tempered, and each industry should develop their own prediction grid to capture the uniqueness of that industry.

Qu et al. [6] determine how to choose the optimal percentage of open-access appointment slots, taking into account provider capacity, no-show/arrival rates and distribution of demand. The success of an open access scheduling system relies on the appropriate percentage of prescheduled and open-access appointment slots. The wrong percentage of open access appointments could result in a mismatch of capacity and patient demand, leading to the failure of the system. The authors aim to find the optimal number of appointments that can be prescheduled, prior to scheduling any appointments for the provider while maximizing the expected number of serviced patients. The formulation and derivation of the quantitative model takes into account expected number of patients that are seen by their physician, patient arrival, provider capacity, appointment demand, and no-show rates. The results demonstrate how the optimal percentage of open access appointments relies on the ratio of provider capacity to average demand for open access appointments. The ratio of patient arrivals for prescheduled and open access appointments is also determined.

Gupta and Wang [15] develop a discrete-time, finite-horizon Markov Decision Process to model patients' choice. The objective is to maximize expected revenue obtainable from periods *t* onwards, given that the clinic's reservation state at time *t* is s. The clinic's reward is based on same-day demand. For a single physician, the model determines a booking limit policy based on the optimal policy, which identifies the number of patient's request that can be accepted. The authors only partially characterize multi-doctor clinics of optimal policy using two heuristics. They also consider the effect of clinic's optimal profit on patient's loyalty to PCP, total clinic load, and load imbalance among physicians. One of the major shortcomings of their research is that they assume a patient can be denied same-day request to protect certain slots for later arriving same-day patients or for patients belonging to the requested physicians' panel.

2.5 Research Gaps

Although most research in overbooking in healthcare consider patients to have homogeneous no-show rates, Zeng et al. [89] develop a clinical scheduling model in which the patients have heterogeneous no-show probabilities. LaGanga and Lawerence [90] developed a simulation model to mitigate the loss of productivity of physicians due to patient no-shows by testing the performance of scheduling rules for overbooked appointments. Muthuraman and Lawley [59] also develop a stochastic overbooking model under an open-access scheduling system for a single service period to compensate for the probability of no-show for outpatient clinics. Liu et al. [91] consider the probability of no-show under a traditional scheduling system, where a Markov decision process is used to determine which day to schedule a patient's request to be seen by their physician. Zeng et al. [92] use a game theoretical approach to model the behavior between the clinic and patients and demonstrate that based on the patient's characteristics overbooking may or may not improve clinic's profit. Moreover, it has been noted that overbooking tends to penalize patients that arrive for their scheduled appointments by increasing the amount of time they spend waiting to see their physician [92-93]. Realizing this dilemma, schedulers must identify other approaches that will not negatively impact patient satisfaction.

Table 3 demonstrates the contributions that have been made towards quantifying the theory and objectives of advanced access scheduling and whether or not revenue management concepts are utilized. The primary goal of this literature review is to understand the models based on the presence of no-show and/or cancelled appointments, the number of providers, the number of patients seeking to schedule a same-day appointment and dependency of the patients amongst themselves. The key to optimizing appointments is to take a quantitative approach to develop the schedule rather than relying on an experts experience. [6] Based on the literature survey, the research seeks to develop a quantitative model to examine the impact of batch (i.e. dependent demand) arrivals under an open-access scheduling system. To date, quantitative papers in this area consider single and multiple provider models under the assumption that demand arrivals are independent among patients. In regards to revenue management, McGill and Van Ryzin [18] stated that it must consider the inclusion of batch bookings as critical area for research.

Table 3. Quantitative model comparison

Author	*Model*	*Single Provider*	*Multiple Provider*	*Dependent Demand?*	*No-Shows?*	*Appointments*	*Distribution*	*Overbooking?*
Qu [6]	Mathematical	X		No	Yes, known	Total number is known and fixed.	Known for prescheduled and OAS	No
Gupta and Wang [15]	Finite MDP and Heuristics	X	X	No		Considers Patient Choice		Yes
Liu et al. [91]	Infinite MDP and Simulation	X		No	Yes	Proposes Improved OAS for traditional system.		No
Kopach [58]	Simulation		X	No	Yes	Allows double-booking	Poisson	No
Giachetti et al. [57]	Simulation	X		No	Yes			Yes
Muthuraman and Lawley [59]	Multiobjective optimization	X		No	Yes	Allows overbooking. Appointment slot allocation. Patient choice.	Exponential	Yes
DeLaurentis et al. [56]	Simulation and Queueing Model		X	No	Yes			Yes

CHAPTER 3

Comparison of Patient Dependency at Public and Private Pediatric Dental Clinics: A Case Study

3.1 Introduction

According to the United States 2009 Census data, households with one, two or three children represent approximately 21%, 38%, and 25% of the living arrangements of children under the age of 18, respectively [94]. In addition, the percentage of households with multiple children under the age of 18 with both parents, a single mother, or a single father present in the households are 84%, 75%, and 64%, respectively. These statistics combined illustrate the potential strain that can be placed on a caregiver to ensure the health needs of the family are attended to (i.e. requiring hours away from work). From the service provider's standpoint, a challenge may arise in keeping flexible schedules to accommodate appointment requests that minimize "time out of work" for the parent. This paper explores the relationship between multi-family appointment requests and patient scheduling.

Since its introduction by Bailey [78], research in the field of appointment scheduling has been extensively studied. In general, appointment systems are designed to minimize waiting times for patients while maximizing the utilization of physicians and other resources [79-81]. There is also a general assumption that patients are independent amongst each other in arrivals and no-show rates within scheduling models [82]. This assumption makes analytical models more tractable, in the sense that knowledge of one patient does not affect the probability that the other patient will arrive. However, this assumption is often invalid in environments where requests for multiple appointments are made within the same family or group. Here, one patient is highly dependent on the interactions of one or more patients within their family or group. Abdus-Salaam et al. [95] introduce the concept of dependent demand arrivals in patient scheduling in which the arrival of patients is dependent among each other within the group; hereafter referred to as a batch. A batch request is defined as at least two individuals who are interested in scheduling appointments with a set of providers if and only if each person within the batch will be served. These patients are scheduled simultaneously or consecutively depending on the number of idle physicians and the number of patients within the batch. We examine this scenario in our study consisting of a public and a private pediatric dental clinic in which parents request an appointment for each of their children within the same time frame.

On average, the clinics studied for this research experience batch appointment requests at nearly 42%. Although the private and public clinics experience similarities, they differ significantly in their no-show and/or cancellation rates. At the private dental clinic, less than two percent of their patients fail to meet their scheduled appointment; whereas nearly twenty percent of the public clinic exhibits the same behavior. The difference in no-show rates is not a surprise. Gupta and Denton [26] present the challenges faced with private and public clinics, in which clinics that predominately serve patients with private insurance or Medicare experience low no-show rates and late cancellations. The authors also recognize that public clinics that serve under/uninsured populations, Medicaid recipients, or patients with mental health issues experience a significant number of no-shows.

The scheduling coordinator at the public clinic acknowledges that patients tend to break their appointments due to a lack of transportation, inclement weather, and scheduling conflicts. Furthermore, 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 [86]. Patients that fail to meet their scheduled appointment can negatively impact the patient's care due to their inability to receive information on how to better manage their health needs [96]. The reason for the appointment, patient's attendance history, appointment session (morning or afternoon), weather, insurance and age group are identified as the key factors in predicting whether a patient will meet their prescheduled appointment [58, 97]. Whether or not a patient is new, is also a contributing factor in determining if a patient will fail to meet their scheduled appointment [98]. It is believed that the single most predictor is based on whether or not a patient attended their previous appointment [99].

This research aims to understand no-show patterns in terms of the number of patients that are scheduled within the same family in both the public and private dental sectors. Although, the scheduling coordinators at the respective dental clinics observe the behavior of parents requesting batch appointments, they have not conducted a detailed analysis on the impact of these appointments on their scheduling paradigm. The clinicians also have not considered if there is a relationship between no-show rates and batch appointments. In addition, the impact of no-shows on the clinics if a family breaks their appointment is much higher as resources are idle during the appointment period.

Using multivariate statistics techniques, the impact of dependent demand arrivals on the dental scheduling system is explored. In addition, these statistical models are developed to understand the prevalence of batch appointment requests and how their prevalence differs based on clinic type. An empirical study is presented to understand the difference between clinic types as it relates to both patient no-show rates and batch appointments. The paper aims to address whether or not there is a relationship among the appointment demand type (batched versus individual), patient behavior (break or meet scheduled appointment), and reason for the appointment.

The remainder of this paper is organized as follows. Section 3.2 provides a brief overview of the clinics background. Section 3.3 distinguishes the difference between clinic types and a discussion of data. Section 3.4 describes the multiway frequency analysis and logistic regression models. Section 3.5 presents the results and analysis of the models. Section 3.6 summarizes and concludes the objectives of this research.

3.2 Clinic Structure and Scheduling Paradigm

The public dental clinic aims to provide exams, treatment, cleanings and emergency care for children. In addition, the clinic provides dental services to pregnant women who have a pink Medicaid card and to uninsured children through the "Prompt-Pay" program. The Prompt-Pay system allows parents to pay at a discounted rate for several services at the time of the appointment. The clinic has nine chairs of which four are hygiene chairs, two are adult chairs, and three are operative chairs. The staff includes two full-time dentists, two dental assistants, and a receptionist. During the summer, a part-time hygienist is available Monday, Wednesday, and Friday. The clinic operates Monday through Friday from 8:00 AM to 4:10 PM. The clinic's daily capacity when there are two dentists and a hygienist available is twenty-six, but without a hygienist only twenty patients can be served.

The two primary appointment types can be categorized as recalls and operative. Recalls consist of general check-up and cleanings, whereas the operative appointments entail cavity fillings and tooth removal. Each type varies in the length of appointment duration. For an operative appointment, the appointment can range from forty to ninety minutes, while recall appointments range from ten to thirty minutes.

The clinic has a first available appointment scheduling policy of which two appointment slots per day are reserved for potential emergency requests. The clinic has experienced challenges with patient scheduling. Typically, patients are doubled-booked to mitigate potential no-show and cancellation of patients and to increase the dentist productivity. Whenever possible, the clinic sends out reminder calls to prevent potential no-shows and cancellations. The clinic does not provide special accommodations for families, but they do allow multiple children to be scheduled within the same time frame. The scheduling coordinator aims to schedule families simultaneously and/or consecutively whenever possible. However, the clinic has implemented a stricter policy for Prompt-Pay families in which families that have a history of breaking their appointment are not allowed to schedule multiple children. In addition, the clinic must prevent potential language barriers for patients by providing an interpreter for non-English speaking patients. On Wednesdays, the clinic has a Spanish interpreter available all day; whereas other patients that require an interpreter are scheduled on an as needed basis.

The private dental clinic provides preventative care and services, restorative dentistry, infant oral care, habit development and management, interceptive orthodontics, trauma treatment and management and emergency treatment for children throughout their community. Similar to the public clinic, the private clinic is able to classify these services as recall and operative appointments. Therefore, recall appointments consist of the preventative care, the infant oral care, and the habit development and management services; whereas, the restorative dentistry, interceptive orthodontics, and trauma treatment and management represent the operative procedures. Unlike the public clinic, each appointment has duration of thirty minutes. However, the clinic does allow longer appointment durations as needed for special needs patients and operative appointments that require extended time.

The staff includes one full-time dentist, three dental assistants, a part-time hygienist, an office manager and an administrative assistant. The clinic's single dentist typically utilizes two chairs to better serve their patients. The clinic has a daily capacity of fifty patients. As noted by the clinic coordinator, patients' demographics consist of 45% Caucasian, 40% African-American, 5% Asian, 5% Hispanic and 5% other ethnicities. In addition, patients have diverse income levels in which payment methods are made by health maintenance organizations (HMO), Medicaid and out-of-pocket (no insurance) nearly 85%, 10% and 5% of the time, respectively.

Currently, the clinic operates under a first available scheduling paradigm, where emergency appointments are worked into the schedule as needed. The clinic allows request of two children per family to be scheduled within the same time frame. In addition, the clinic aims to schedule younger children early in the day; whereas, older children are seen outside regular school hours whenever possible. In regards to no-shows and cancellations, the clinic has a 24 hour notice for cancellation policy or else the patient must adhere to a $25 broken appointment fee. The clinic does not allow the use of overtime to accommodate the service of additional patients. The clinic hours of operation are as follows: Monday, Wednesday and Thursday from 7:45 AM to 3:30 PM, Tuesday from 8:00 AM to 4:30 PM and Friday from 8:00 AM to 12 Noon for administrative hours only.

3.3 Analysis Method

First, this research compares and contrasts the differences in the public and private pediatric dental clinic. In addition, this work seeks to investigate the following: What is the throughput of each clinic type in terms of appointment demand type? What is the primary reason that patients request an appointment for each demand type? Does the day of the week contribute to whether or not an individual and/or batch meet their scheduled appointment? What is the probability of no-show for each clinic in terms of the number of appointments requested? Are patients that are scheduled within the batch truly dependent amongst each other? What is the financial impact of scheduling families for the respective clinics? Parents/caretakers?

Upon the approval from the Institutional Review Board (IRB), scheduled and broken appointments data are obtained from the dental clinics from April 1, 2009 to September 30, 2009. Table 4 displays the number of entries and key fields for both clinic types. With the broken appointments data, the patients are classified based on whether or not they failed to meet their scheduled appointment without notice or calls to cancel. Each clinic provided information regarding appointment date, the reason, and the telephone number of the parent(s) or caretaker. The public dental clinic also provided data on the provider (since there are multiple dentists) and duration of the appointment. For the private clinic, at most only 2052 out of the 2090 entries are used for our data and statistical analysis. This is a result of either scheduling inconsistencies or a single patient being scheduled for multiple appointments on the same day. In the latter instances, an entry is deleted and noted in the reason for the appointment for the remaining entry (if there are multiple appointment slots being allocated to a single patient). It is worth noting that the appointment schedule does not indicate the time the actual appointment request is made; nor does the broken data identify when a patient calls to cancel.

In order to determine which patients belonged to a family group, there is a general assumption that on a given day that each patient with the same last name and/or telephone number belonged within the same appointment group. To validate this assumption, the appointment time is used to verify if patients are scheduled within a batch appointment request. In addition, the scheduling coordinator for the respective is contacted to confirm whether or not patients were indeed a family. For example, if two patients have the same last name but different telephone numbers and the appointments are not scheduled within a two hour time window, then we assumed the patients may not be a part of a batch appointment request. These assumptions are essential in identifying how many individuals are scheduled within the same family, since batched patients cannot be determined solely by relying on just the patient's last name. Especially, in the event that a family has several children each having different last names.

Table 4. Scheduling data set

	Scheduled		**Broken**	
Clinic Type	*Entries*	*Key Fields*	*Entries*	*Key Fields*
Public	1246	Appointment Date	280	Appointment Date
		Patient's Name		Patient's Name
		Provider		Provider
		Reason		Reason
		Duration		Duration
		Phone Number		Phone Number
				Status
Private	2048	Appointment Date	42	Broken Date
		Appointment Time		Patient's Name
		Provider/Room		Provider
		Reason		Reason
		Patient's Name		Duration
		Phone Number		Phone Number
				Status

Figure 8 provides a sample schedule in which there are three families scheduled consecutively throughout the day. Here, four children from the same family (i.e the Jeffersons) are scheduled consecutively within a two hour period. In the event that these patients failed to meet their scheduled appointment, the clinic will suffer an immediate decrease in productivity. The figure also suggests that the clinic aims to schedule families that are scheduled consecutively within the same room whenever possible. However, the Fox children are scheduled in separate rooms, but within an hour and a half time frame. Note *: to protect the patient's identity fictitious names are used.

				Friday, August 21, 2009		
Time		Op-1	Op-2	Op-3	Op-4	Op-5
7am	:15	N. Taylor Recall				
	:30					
8am	:00	K. Ovens Recall	Z. Marsh Recall	S. Fox Recall	A. Washington Oper	
	:30	K. Walker Recall				
	:45				M. Cambell Oper	
9am	:00	M. Brown Recall	A. Favors Recall			
	:30	C. Fox Recall			C. James Oper	C. Baker Emerg
	:45		B. Favors Recall	F. Jefferson Recall		
10am	:00					K. Couch Oper
	:15			I. Jefferson Recall	E. Griggs Emerg	
	:30	S. Mack Recall	S. White Recall			
	:45			J. Jefferson Recall		M. Robinson Oper
11am	:00	J. King Recall			I. Thomas Oper	
	:15			G. Jefferson Recall		
	:30	K. Camden Recall	D. Johnson Recall		G. Drake Oper	
	:45			J. Conner Emerg		
12pm	:00	L. Jackson Recall	A. Moore Recall			A. Humphrey Oper
	:15				B. Arthur Oper	
Legend:		Batch Size=2	Batch Size=4			

Figure 8. Sample schedule of consecutive batch appointment scheduling at the private clinic*

Figure 9 illustrates the scenario where several families are scheduled both consecutively and simultaneously. The figure also demonstrates the importance of validating whether or not patients are truly scheduled within a batch. In addition, we observe cases in which a patient is scheduled twice within the same period (i.e. M. Bradley) and when a patient is scheduled for multiple appointments (i.e. H. Bishop). Although not illustrated within the figures, the public clinic exhibits similar scheduling complexities.

Tuesday, April 14, 2009

Time		Op-1	Op-2	Op-3	Op-4	Op-5
7am	:15					
	:30					
8am	:15	O. Graham Recall	M. Glover Recall		J. Thompson-Foy Oper	
	:30					K. Foy Oper
	:45	N. Graham Recall				
9am	:00					
	:30	J. Mills Recall**				
	:45					
10am	:00		H. Bishop Recall		M. Bradley Recall	M. Bradley Recall
	:15					
	:30	M. Jeferrson Recall	S. Bishop Recall		M. Spencer Oper	
	:45					
11am	:00	K. Dinkins Recall**	J. Haynes Recall		H. Bishop Oper	
	:15					
	:30	A. Primus Recall	D. Payne Recall			
	:45					
12pm	:00	C. Primus Recall				
	:15					
1pm	:00			LUNCH		
	:30					
2pm	:00	G. Jessamy Recall	O. Jessamy Recall		C. Mills Oper**	
	:30	M. Jessamy Recall	A. Jessamy Recall			
3pm	:00	M. Ealey Recall		E. Austin Recall		
	:30	T. Ealey Recall			S. Francis DA	
4pm	:00	T. Simpson Recall	H. Dinkins Recall**		M. Shealy Recall	S. Moffitt DA
	:30					
Legend:		Multiple Appointments	Batch Size=2	Batch Size=4	**Verify if family	

Figure 9. Sample schedule of simultaneous and consecutive batch appointment scheduling at the private clinic*

Table 5 displays the actual parameters that are used for the statistical model along with their respective levels and values. Note: the asterisks indicate that the provider and duration data are available only for the public clinic. The latter is based on the fact that the private clinic typically only allocates thirty minutes to appointment requests. In terms of the provider parameter, the focus of our analysis is directed towards the dentist value.

Table 5. Variables

Parameter	Levels	Values
Day	5	Monday, Tuesday, Wednesday, Thursday, Friday
Patient Behavior (Status)	2	Scheduled or Broken
Provider*	3	Dentist, Hygienist, Ortho
Reason	3	Recall, Operative, Emergency
Duration (in minutes)*	6	≤20, 30, 40, 50, 60, >60
Demand Type (Batch)	2	Yes or No (Batch or Individual)

3.3.1 Data analysis

3.3.1.1 Aggregated analysis

During the six months period, the total number of appointments for the private and public clinics is 2053 and 1526, respectively. Figures 10 and 11 display the actual number of appointments for each day over the observed period. For Figure 10, the maximum, minimum, and average number of patients seen at the public clinic is 30, 1, and 12, respectively. The clinic served batched patients at a maximum, minimum, and average of 15, 2, and 5.7, respectively. The maximum number of batch requests occurred on the day in which the Spanish interpreter was available. This suggests that public clinic aims to serve as many families as possible when additional resources are needed. Recall, that the scheduling coordinator originally stated that the public clinic has a daily capacity of 26. In order to achieve the higher observed maximum capacity value, the dentist available for that day served two patients per appointment block. As evident in Figure 11, the private clinic encountered a maximum of 40, a minimum of 2, and an average of 20 appointments per day in regards to the total number of patients that are served. This suggests that the private clinic never reaches its full capacity throughout our period of interest. In addition, the maximum, minimum, and average number of batched patients scheduled is 24, 2, and 9.9, respectively. On average, the private clinic schedules 58% more families than those scheduled by the public clinic.

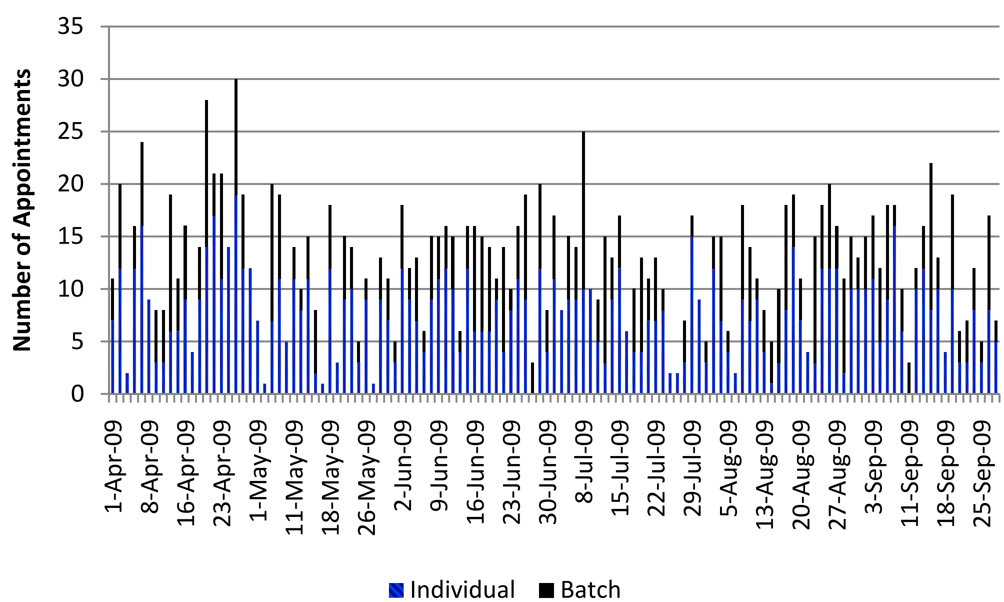


Figure 10. Individual and batch appointments for public clinic

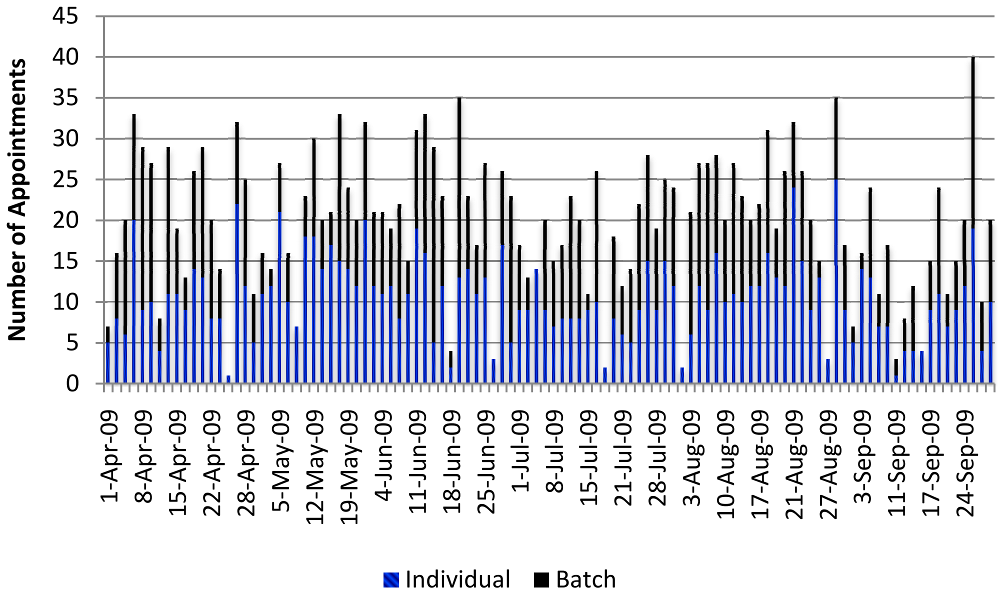


Figure 11. Individual and batch appointments for private clinic

Figure 12 illustrates the relative frequency of each variable, which is calculated as a function of the clinic type. The low frequency of appointments on Friday is contributed to the public clinic donating its facility to the orthodontics program and the private clinic is typically closed. Figure 12(a) shows that nearly 28% of private appointments are on Mondays and Tuesdays, whereas the public clinic experiences the most appointments at approximately 25% on Wednesdays. In regards to the day of the week, the public clinic tends to schedule patients that require a Spanish interpreter (if their primary language is not English) on Wednesdays. As evident in Figure 12(b), 98% of private patients meet their scheduled appointments; whereas only 82% of public patients exhibit the same behavior.

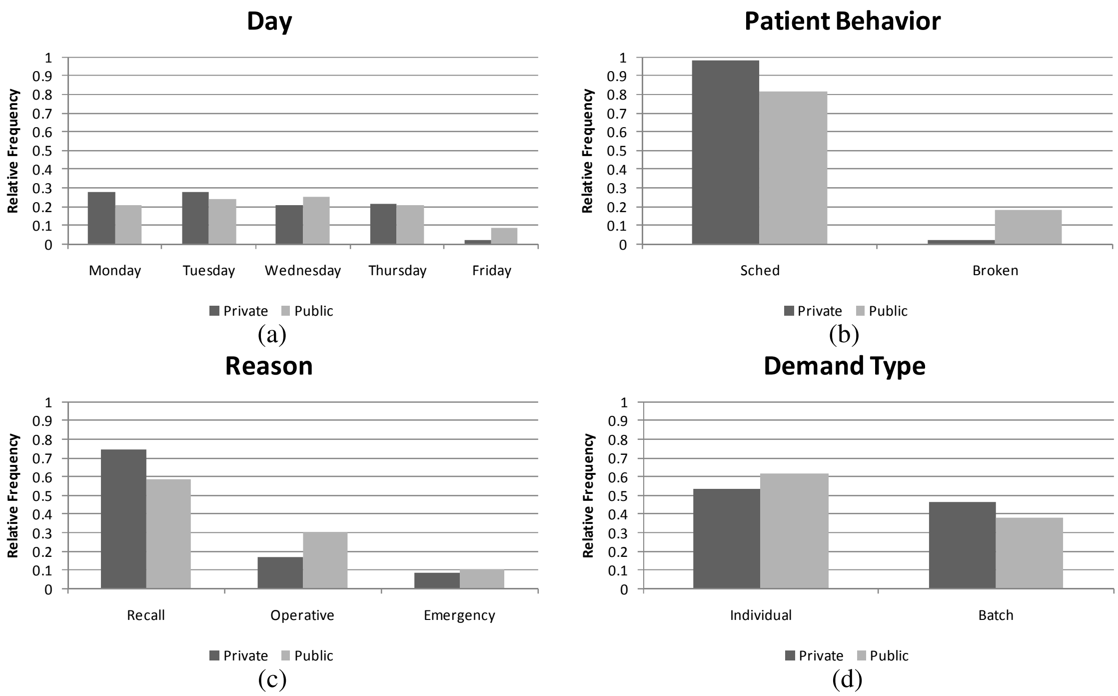
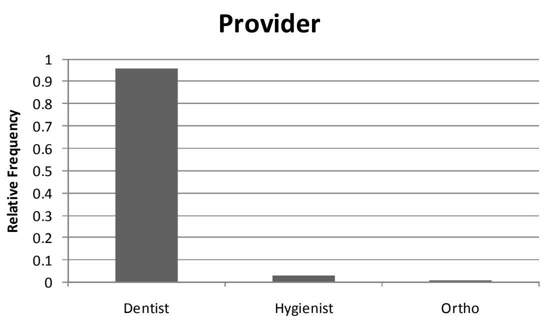


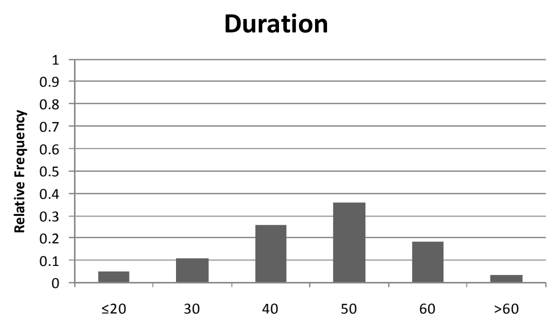
Figure 12. Clinical comparison for day (a), patient behavior (b), reason (c), and demand type (d)

Figure 12(c) shows that both clinics primarily schedule patients for recall appointments. However, the number of operative and emergency appointments is significantly higher at the public clinic. This implies that the private clinic's patients often require preventative care; whereas the public clinic's patients require restorative treatment and care. In addition, the scheduling coordinator suggests that the large number of emergency appointments is due to the dentist not fully meeting all of the patient's needs at the time of their original appointment.

Furthermore, both clinics encountered batched appointments at 46% for private and 38% for public as seen in Figure 12(d). However, the largest batch size is four for private and six for public. The most frequently requested batch size for both clinics is two at approximately 24% for the public and 38% for the private. This coincides with the fact that families with two children are the highest occurring family size in the United States. This suggests that the scheduling coordinators for each of the respective clinics consider the allocation of appointments for at least two patients when determining appointment times.



(a)



(b)

Figure 13. Public clinic additional relative frequencies for provider (a) and duration (b)

Figure 13 displays the additional parameters obtained from the public clinic. In Figure 13(a), the dentist provides care to the patients about 96% of the time. As illustrated in Figure 13(b), nearly 80% of the appointments are between forty to sixty minutes. Some important facts about the public clinic can explain the variations in the relative frequency across the respective parameters. The clinic is faced with the challenge of being understaffed. Therefore, the available dentist provides the majority of the service on a given day, since there is not a full-time hygienist. Also, the clinic has been reduced from two dentists to a single dentist in the last four months of the data period. During this period, we observe that average number of patients scheduled has been reduced slightly from 12.7 with two dentists to 12.2 with only a single dentist. However within this same period, the average number of patients scheduled within a family has slightly increased from 5.6 with two dentists to 5.7 with only a single dentist. Thus, the clinic remained productive in spite of the reduction in personnel.

3.3.1.2 Detailed analysis

To further understand the prevalence of batch appointments at each clinic, we illustrate the relative frequency of batch and individual appointments for both the day of the week and patient's behavior in Figure 14. For the public clinic, Figure 14(a) show that patients scheduled with a batch or as an individual meet their prescheduled appointment the most on Tuesdays and Wednesdays, respectively. As with Figure 14(b), nearly 70% of the appointments are broken from Monday through Wednesday. Of these broken appointments, batched appointments are the highest on Wednesdays at 25%. In addition, the highest percentage of batch requests in which patients meet their scheduled appointments is on Tuesdays at 26%. As with the private clinic, Figure 14(c) show that patients meet their scheduled appointment scheduled the most on Tuesdays for batched patients and on Mondays for individual. In Figure 14(d), no appointments are broken on Fridays. In fact, the highest observed broken appointments are on Mondays for batch requests and Thursdays for individual requests.

Although not illustrated by the figure, the impact of the batch size on the clinics' appointments is also examined. At the public clinic, only the families requiring three or less appointments are booked on Fridays. In addition, families with five or more children are scheduled on Tuesdays through Thursdays. Therefore, it is likely that these families required the use of an interpreter. For the private clinic, family sizes of two and three are the highest on Tuesdays. Mondays experienced the highest number of individual appointments. Other than Fridays, Wednesdays observed the least amount of batch and individual appointments. Based on these results, there are no general assumptions that can be made in terms of the scheduling of individuals versus batched patients at either clinics. This also suggests that parents/caregivers request batch appointments on an as needed basis.

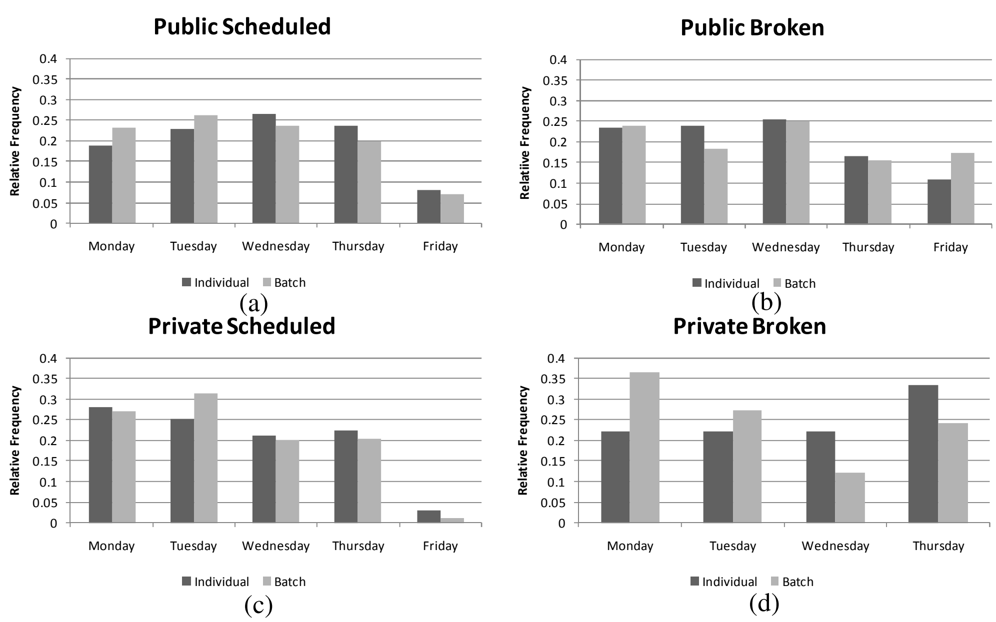


Figure 14. Relative frequency of scheduled and broken appointments for each clinic type

3.3.2 Batch size and probability of no-show

3.3.2.1 Probability of no-show

Table 6 provides the probability of no-show for the respective appointment request size values for the public and private clinic. Once the number of patients per family are determined, the probability of no-show values are calculated based on the data provided from each clinic. The no-show rate is derived from the frequency and the total number of patients that fail to meet their scheduled appointment for the respective request size values. These values represent the probability of the entire batch not meeting their scheduled appointment. As a reminder, each family size is identified by their last names and/or telephone number on a given day.

From the table, the probability of no-show is higher for both clinics when two children are scheduled within a batch. Also, at the higher observed batch size values for the respective clinics, there is a guarantee that every patient within the batch will meet their scheduled appointment. It is worth noting that the private clinic does not typically allow batch sizes greater than two. However, there are special situations in which a larger batch size will be accepted, i.e. family history. Based on data from both clinics, there are only three occurrences out of the 693 batch appointment requests that do not result in patients being dependent in arrival. Therefore, each patient within the batch request will be dependent upon the others. Thus, if one patient fails to meet their scheduled appointment, then the entire batch will fail to meet their scheduled appointment.

Table 6. Frequency and probability of no-show for given batch size values

	Public			Private		
Batch Size	*Frequency*	*P(no-show)*	*P(show)*	*Frequency*	*P(no-show)*	*P(show)*
1	941	0.1870	0.8130	1100	0.0082	0.9918
2	185	0.2081	0.7919	388	0.0387	0.9613
3	43	0.1163	0.8837	52	0.0192	0.9808
4	16	0.1875	0.8125	5	0.0000	1.0000
5	2	0.0000	1.0000	0	-	-
6	2	0.0000	1.0000	0	-	-

3.3.2.2 Financial impact of demand type

The final analysis is to determine the impact of the appointment demand type on the clinic's total profit and the cost of inconvenience for parents being out of work. Similar to Moore at el. [98], the loss of revenue due to patients failing to meet their scheduled appointment is estimated. Table 7 summarizes the total profit for each clinic type and appointment demand type. The revenue (R) generated for serving a patient is the same for each clinic type at $135 per patient. Equation 3.1 calculates the total profit over the data period (T) as a function of both the loss of revenue due to broken appointments (N) and the revenue generated from serving patients for the respective demand types (S). The total profit for the batch appointments are observed from the smallest to the largest possible request size (B) for each of the respective clinics. For the individual demand type, the equation is further simplified to accommodate only the request size of one.

$$\text{Profit} = \sum_{t=1}^{T} R * \left[\sum_{b=2}^{B} (S_{b_t} - N_{b_t}) \right] \tag{3.1}$$

From the table, for the private clinic, the loss of revenue due to batch appointments (79%) is significantly higher than those of individual appointment requests (21%). As with the public clinic, individuals account for roughly 63% of the total loss of revenue yielded as a result of broken appointments. In addition, the private clinic generated approximately 45% of their total profit from batch appointment requests; whereas, the public clinic experienced nearly 39%. The patients that require a single appointment (for public) and the patients that are booked as a group (for private) yielded the highest ratio of loss of revenue over generated revenue. These results coincide with those expressed in Figure 14 in regards to the number of patients that fail to meet their scheduled appointment based on both clinic type and demand type.

Table 7. Impact of demand type on profit

Clinic Type	Demand Type	Lost Revenue	Revenue	Total Profit
Public	Batch	\$14,040	\$64,935	\$50,895
	Individual	\$23,760	\$103,275	\$79,515
	Overall	\$37,800	\$168,210	\$130,410
Private	Batch	\$4,455	\$124,065	\$119,610
	Individual	\$1,215	\$147,285	\$146,070
	Overall	\$5,670	\$271,350	\$265,680

In order to quantify the cost of inconvenience to parents in scheduling multiple children, both the loss of income due to absence from work and the amount of time needed to serve their appointment request are considered. Table 8 displays the effects of parents having to schedule multiple children given that the patients are served simultaneously, both simultaneously and consecutively, consecutively with a gap (waiting time), and consecutively without a gap. The actual cost of inconvenience is based on the hourly income of the households in the United States, which is derived from the Census 2008 data on the median income for families. Note; the scheduling pattern is based on the sample schedules previously shown in Figures 8 and 9. There is a significant difference for patients that are scheduled consecutively with a gap. For example, the parents with the longer waiting time between their children being served experiences a 40% increase in the cost of inconvenience given each request size is two. The longer the waiting time, the more likely the parent will leave and return for the latter appointment. However, it is essential to understand not only the time out of work factor, but also the leaving-and-returning factor. Thus, the clinics should aim to balance both the impact of demand type on their profit and the inconvenience of parents having to be out of work.

Table 8. Cost of inconvenience to parents

Schedule Pattern	Batch Size	Duration	Cost of Inconvenience
Simultaneously	2	0:30	$12.54
Simultaneously & Consecutively	4	1:00	$25.08
Consecutively with gap	2	2:00	$50.17
	2	5:00	$125.42
Consecutively without gap	2	1:00	$25.08
	4	2:00	$50.17

3.4 Statistical Method

To determine if there is relationship among the parameters listed in Table 9, two multivariate statistical models are generated. Given the characteristics of each parameter, we determined that multiway frequency analysis and logistic regression will best fit our data. These models are utilized when there are multiple discrete independent variables and a single dependent (if any) variable. The multiway frequency analysis is used to identify the degree of relationship among variables [100].

Table 9. Frequency for the public and private pediatric dental clinics

		Public Clinic			***Private Clinic***		
		Reason			**Reason**		
Demand Type	**Patient Behavior**	*Recall*	*Operative*	Total Frequency	*Recall*	*Operative*	Total Frequency
Batch	*Scheduled*	391	86	477	822	66	888
	Broken	83	17	100	33	0	33
Batch Total		474	103	577	855	66	921
Individual	*Scheduled*	311	306	617	633	275	908
	Broken	98	54	152	5	2	7
Individual Total		409	360	769	638	277	915
Total Frequency		883	463	1346	1493	343	1836

The goal of this work is to determine if there is a relationship among appointment demand type, patient behavior, and reason for the appointment. The appointment demand type is used to identify whether or not a patient is within a batch; whereas patient behavior is based on whether or not a patient met their scheduled appointment. The null hypothesis for the full effect model states there is no relationship among appointment demand type, patient behavior and reason for appointment. The respective frequency values for each possible case for the appointment demand type, patient behavior, and reason variables are provided in Table 9. The model is used to examine the association of each variable from the one-way, two-way and higher order frequency table. The model is tested using the Chi-squared (χ^2) test of significance.

In order to predict group membership, logistic regression enables one to create a linear combination of the log of the odds of being in one group. This model is constructed to determine which variables contribute to the probability of patients meeting their scheduled appointment. Thus, this research addresses the following: Can the patient behavior be determined based on the day, provider*, reason, duration* and appointment demand type? A statistical stepwise regression approach is used, since this research is data-driven. The model is tested using the Chi-squared (χ^2) test of significance with an alpha of 0.05. [100]

3.5 Results

3.5.1 Multiway frequency analysis

For the public pediatric dental clinic, the associations of characteristics are identified for a sample size of 1346 and a response level of eight. Tables 10 and 11 display the results of the analysis of maximum likelihood estimates for the public clinic for the full, second, and first order effects. Using the maximum likelihood analysis of variance for the full effect, there is no significant association among appointment demand type (batch or individual), patient behavior (scheduled or broken) and reason for appointment. This implies that given the fact that a patient is scheduled within a batch and the reason for the appointment is known, the scheduling coordinator will be unable to determine whether or not the patient fails to meet their scheduled appointment. Therefore, we fail to reject our null hypothesis. However, the second order effect provides the dominant interaction. With the second order, there is a relationship between appointment demand type and the reason for the appointment. Requests for operative appointments are made by individuals nearly 78%, whereas 54% of recall appointments are made by batched patients. Based on the first order effect, appointment demand type, patient behavior and reason for appointment are proven to be statistically significant. With respect to the reason for appointment, operative appointments account for 34%. Broken appointments due to cancellation or not showing up are slightly high at 19%. In relation to appointment demand type, those patients that are booked as individuals represent 57% of requested appointments.

Table 10. Higher order effects analysis of maximum likelihood estimates for public clinic

Parameter	Levels	Estimate	Standard Error	Chi-Square	Pr > ChiSq
type*behavior*reason	individual*broken*operative	-0.0636	0.0433	2.15	0.1426
type*behavior	individual*broken	0.0352	0.0433	0.66	0.4168
type*reason	individual*operative	0.3110	0.0433	51.46	<.0001
behavior*reason	broken*operative	-0.0814	0.0433	3.52	0.0605

Table 11. First order effects analysis of maximum likelihood estimates for public clinic

Parameter	Levels	Estimate	Standard Error	Chi-Square	Pr > ChiSq
type	individual	0.2953	0.0433	46.40	<.0001
behavior	broken	-0.7576	0.0433	305.40	<.0001
reason	operative	-0.4640	0.0433	114.58	<.0001

Table 12 displays the results of the higher order effects analysis of maximum likelihood estimates for the private clinic. Similarly, the associations of characteristics are identified for a sample size of 1836 and a response level of seven for the private pediatric dental clinic. Again, there is no significant association among appointment demand type, patient behavior and reason for appointment based on the maximum likelihood analysis of variance for the full effect order. Therefore, we fail to reject our null hypothesis for the full effect model. The second order effect provides the dominant interaction, in which there is a relationship between appointment demand type and the reason for the appointment. In fact, operative appointments for individuals account for 81%; whereas, 53% of recall appointments are for patients that are booked as a batch. In addition, there is a relationship between appointment demand type and patient behavior. As a result, we determined that individual patients rarely (0.8%) break their scheduled appointments. In fact, patients that are scheduled within a batch account for nearly 83% of all broken appointments.

Table 13 displays the results of the analysis of maximum likelihood estimates for the first order effects. With the first order effect, patient behavior and reason for appointment are significant, since the test statistic is less than the significance value. Operative and recall appointments account for 19% and 81%, respectively. Unlike with the public clinic, broken appointments due to cancellation or not showing up are fairly low at 2%.

Table 12. Higher order effects analysis of maximum likelihood estimates for private clinic

Parameter	Levels	Estimate	Standard Error	Chi-Square	Pr > ChiSq
type*behavior*reason	individual*broken*operative	.	.	.	.
type*behavior	individual*broken	-0.4064	0.1207	11.34	0.0008
type*reason	individual*operative	0.4221	0.0367	132.07	<.0001
behavior*reason	broken*operative	-0.0206	0.2096	0.01	0.9215

Table 13. First order effects analysis of maximum likelihood estimates for private clinic

Parameter	Levels	Estimate	Standard Error	Chi-Square	Pr > ChiSq
type	individual	-0.1150	0.1248	0.85	0.3567
behavior	broken	-2.0347	0.1825	124.30	<.0001
reason	operative	-0.8596	0.2113	16.55	<.0001

3.5.2 Logistic regression

For the public pediatric dental clinic, patient behavior is assigned as the dependent variable; whereas day, reason, provider, duration, appointment demand type, and batch size are set as the independent variables. The number of observations read and used is 1526 for the clinic. The number of appointments that are broken due to cancellation and not showing up are 49 and 231, respectively. The other 1246 observations represent those patients that met their scheduled appointment. The model is constructed to determine the probability of having a broken appointment. Using the stepwise selection, the logistic regression equation is simplified to the important parameters that contribute to the model. From table 14, the day, provider, reason, duration and batch size variables contribute to the probability of a patient not meeting their scheduled appointment. This excludes only the appointment demand type variable as a contributing factor, since the percentage of patients requiring a batch appointment is approximately 40%.

Using the estimates from the table, the following logistic regression equation is generated.

$$\ln\left(\frac{p}{1-p}\right) = 2.2516 + 0.3229m - 4.4670d - 3.1178h + 0.3342r + 0.6257eq50 + 0.6215eq60 - 0.6329s3 \qquad (3.2)$$

By solving equation 3.2 with respect to *p*, the probability of no-show can be computed. For example, if a parent requests three recall appointments for their children with a dentist that lasts fifty minutes on a Monday, the probability of no-show is 0.17. In addition, if the same appointment request is made for a single patient or any other batch size value, then the probability of no-show is 0.28. The probability of no-show is slightly lower at 0.17, when there is a request for three recall appointments with the dentists that lasts sixty minutes on a Monday. Thus, the longer duration, the less likely the patient will fail to meet their scheduled appointment. The model also suggests that if the appointment is for any other reason for a family of three, then the probability of no-show is reduced to 0.13. Moreover, the probability of no-show for a single patient or any other batch size can be decreased significantly to 0.13, if the patient(s) is scheduled for a recall appointment that is less than fifty minutes on any day other than Mondays. Under similar conditions, the probability of no-show will be further reduced if the request is for an operative or emergency appointment at 0.09.

Table 14. Analysis of maximum likelihood estimates for public clinic

Parameter	Levels		DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
	Intercept	3	1	0.1081	0.4831	0.0500	0.8230
	Intercept	2	1	2.2516	0.5004	20.2479	<.0001
Day	Monday *(m)*		1	0.3229	0.1611	4.0158	0.0451
Provider	Dentist *(d)*		1	-4.4670	0.5130	75.8323	<.0001
	Hygienist *(h)*		1	-3.1178	0.5968	27.2961	<.0001
Reason	Recall *(r)*		1	0.3342	0.1461	5.2309	0.0222
Duration	Equal50 *(eq50)*		1	0.6257	0.1577	15.7401	<.0001
	Equal60 *(eq60)*		1	0.6215	0.1895	10.7542	0.0010
Batch Size	Size3 (s3)		1	-0.6329	0.2902	4.7574	0.0292

For the private pediatric dental clinic, the number of observations read and used is 2052 when patient behavior is assigned as the dependent variable. Here, the independent variables are the day, reason, appointment demand type, and batch size parameters. The number of appointments that are broken due to cancellation and/or not showing up is 42. The other 2010 observations represent those patients that met their scheduled appointment. Based on the results presented in Table 15, only the batch size variable contributes to the probability of a patient not meeting their scheduled appointment. Again, this excludes the day, reason for the appointment, and appointment demand type variables.

Table 15. Analysis of maximum likelihood estimates for private clinic

Parameter	Level	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
	Intercept	1	-4.6571	0.2900	257.8187	<.0001
Batch Size	Size2 *(s2)*	1	1.4436	0.3447	17.5422	<.0001

The following logistic regression equation is generated based on the estimates provided in the table.

$$\ln\left(\frac{p}{1-p}\right) = -4.6571 + 1.4436\,s2 \tag{3.3}$$

Given that a parent requests two appointments for their children at the private clinic, the probability that the family will not meet their scheduled appointment is 0.04. However, if a parent only request an appointment for one child, the probability that the patient will fail to meet their scheduled appointment is 0.01. Therefore, as expected the probability of no-show is higher for those appointments that are made within a batch request size of two.

For both clinic types, the logistic regression model that the batch request size contributes to the probability of no-show; whereas, the actual appointment demand type does not. However, the multiway frequency analysis suggests that there is a relationship between appointment demand type and patient behavior (i.e. no-show) for the private clinic. This implies that on a disaggregated level knowing the patient's demand type is not as sufficient as knowing the number of people scheduled within an appointment request. Although only the batch request size proves to be significant for the private clinic, we recognize that more information is available from the public clinic. With this additional information, the regression model becomes more complex; which in turn, leads to a better predictor in whether or not a patient will not meet their scheduled appointment.

In addition, the probability of no-show is significantly smaller for the private versus the public clinic which yields a smaller model for the private clinic. The smaller probability of no-show for the private clinic also generated a weaker model for the multiway frequency analysis model. In fact, one of the limitations of the multiway frequency analysis model is that the expected cell frequencies for all of the two-way associations should be greater than one and more than 20% are less than five [100]. To combat this issue, the emergency request level is eliminated as a reason for the appointment; in turn, a slightly better model for both clinic types is generated.

3.6 Conclusion

The intent of this chapter was to analyze the prevalence of batch appointments and no-shows at both a public and private pediatric dental clinic using multivariate statistics. First, the clinics studied for this research experience batch appointment requests at nearly 42%. In fact, the data from both clinics supported the initial claim that each patient within the batch request will be dependent upon the others. Also, the overall patients at the private clinic had a significantly lower no-show rate at 2%; whereas the public experienced a no-show rate of 18%. Thus, this research demonstrated, through the empirical analysis, the sentiment of Gupta and Denton [26] in that there is a significant difference in no-show rates for those clinics that predominately serve Medicaid patients and those who serve patients with private insurance. Next, this work identified (for both clinics) if there is a relationship among the appointment demand type, patient behavior, and reason for the appointment.

As a result of the full effect model developed using multiway frequency analysis, there is no significant relationship among the appointment demand type, patient behavior, and reason for the appointment variables. Based on the second order effect, each clinic experienced that operative requests are significantly higher for individual rather than batch appointments. In addition, similar to the literature; private clinics have significantly lower probability of no-show rates. This is believed to be a result of the economic status of patients, i.e. Medicaid versus private insurers.

Based on the logistic regression model, equations were generated to determine which variables contribute to the probability of patients meeting their scheduled appointment. For the public clinic, the day, provider, reason, duration and batch size variables contribute to the probability of a patient meeting their scheduled appointment. In addition, only the batch size variable contributes to the probability of no-show for the private clinic. Therefore, it is not necessarily if the patient is scheduled within in the batch, but how many patients that are scheduled within the family or group.

Finally, the results of both models will assist the scheduling coordinators at each clinic in determining which batch appointment requests to accept given the reason for the appointment and probability of no-show. Based on the clinical data, the probability of no-show is higher for those patients requesting batch appointments versus individual requests. As a result, this research suggests that clinics understand the history of each family not meeting their scheduled appointment prior to accepting their request for multiple appointments. This will help the clinics manage the risk of scheduling batch appointments. Furthermore, the public clinic should consider only allowing no more than three children per family to be scheduled within the same time frame to mitigate any risk associated with unutilized appointment slots. In the future, the impact of the patient's demographics on no-show rates and batch appointment requests should be explored. The demographics can consist of the patient's ethnicity, income level, insurance provider, family size, single parent or both parents, etc. By adding these demographic indicators, a further investigation of the difference between clinic types can be determined.

CHAPTER 4

Influence of Batch Appointments in Clinic Scheduling: The Infinite-Horizon Case

4.1 Introduction

With surging healthcare cost (approximately $2.5 trillion in 2009 or 17.6% of the nation's Gross Domestic Product) [1], the United States is currently identifying methods to reform the existing healthcare system. In order to combat rising cost, healthcare administrators must determine ways to manage the daily supply of resources with growing demand. One of the most common approaches to handling this problem is to use patient scheduling to balance supply and demand. In general, appointment systems are designed to minimize waiting times for patients while maximizing the utilization of physicians and other resources [79-81].

Typically, research in patient scheduling considers single provider models and assumes patient appointment requests and arrivals are independent amongst each other. However, it is possible for appointment requests to be dependent in the sense that the interactions of one patient are dependent on at least one other patient, especially when requests are for members in the same family. This is particularly true of clinics whose primary patient demographic consists of children. This behavior is observed at a local Medicaid pediatric dental clinic, where parents often request multiple appointments for their children (i.e. batch appointment request). These requests are typically for appointment slots that accommodate simultaneous or consecutive scheduling patterns. While many of the requests were for two children, there were instances of scheduling three or more children either consecutively or sequentially [101]. Since there is some evidence that appointment requests and scheduling can be dependent, it also suggests that no-show rates between patients may also be dependent. In the case of pediatric clinic scheduling, the entire family could break their appointment. In a clinic offering public health services for children, such factors as lack of transportation, inclement weather, and scheduling conflicts could account for this behavior. The reader is referred to Abdus-Salaam and Davis [101] for a detailed case study summarizing the influences of dependent demand arrivals.

Given the risk associated with the acceptance and scheduling of families, healthcare providers must find ways to balance the needs of the patients in a manner that does not reduce the physicians' utilization and the clinic's profitability. Therefore, we explore the use of open-access scheduling systems. Open-access, also referred to as advanced access or same-day appointment scheduling, transforms traditional scheduling systems into prescheduled and same-day appointments. This system shifts to a patient-centered model which aims to provide timely access to care and improve continuity of care, while allowing patients to see their primary physician within the same-day of the request for an appointment [6, 8]. As a result, patients are able to be seen at their leisure, in turn, improving healthcare delivery quality while reducing healthcare cost. However, there is a fundamental challenge in identifying methods to manage capacity, while meeting daily patient demand [9]. Therefore, advanced access scheduling limits the amount of prescheduled appointments for a specific time frame. There is also an associated risk in managing no-shows with prescheduled appointment slots and undermining the capability of open-access appointments with too few open appointments available [6]. Much of the open-access literature seeks to understand the framework and objectives of open-access scheduling [10, 22-23, 25, 27-28, 30, 32], while others discuss the process of implementing the scheduling system in healthcare clinics [34-43, 102]. Additional literature focuses on the outcomes of implementing the scheduling system in a variety of healthcare clinics with financial, provider satisfaction or statistical analysis [9, 44-55]. Although little work has been done on the quantitative aspect, several papers attempt to provide logical reasoning to support the concept of open-access scheduling with a variety of mathematical models [6, 15, 57-60, 103]. The objective of this research is to present a quantitative method that identifies scheduling rules that are ideal for the acceptance of batch appointments under an open-access scheduling system.

Motivated by the work of [101], a theoretical model is presented to study the impact of dependent demand arrivals on an open-access scheduling system. This study is framed around the following research questions: How is the open-access scheduling paradigm, in terms of the percentage of the appointment slots that are allocated to same-day request, affected by the batch appointment requests? How is the optimal scheduling policy affected by varying degrees of the percentage of prescheduled patients, patient behavior, and appointment request size? How does overbooking affect the performance of the clinic? Given the influx of demand at a single point in time (i.e. batch request), this work also examines if the use of overbooking increases the acceptance of batch appointments. To address these questions, a stationary, discrete time, infinite horizon Markov decision process is presented to model the dynamics of the clinic in the long-run.

The infinite horizon model allows us to identify which scheduling rules should be implemented for each possible state, regardless of the timing of appointment request. Therefore, the underlying assumption is that the decision is stationary for a specific number of patients that are prescheduled and the number of patients that are overbooked. The behavior of the system is quantified under several performance measures including the total expected number of patients that are served, the utilization of physicians, and the expected number of patients assigned to the backlog (i.e. overbooked). The backlog represents the queue of patients that are waiting in the system to be seen by the first available physician. The results indicate that the model tends to always accept a request for batch appointments when the probability of no-show is greater than or equal to 0.5. Also, with the acceptance of batch requests, the total expected number of patients that are served decreases as the probability of no-show increases. In addition, the expected number of patients that are waiting in the backlog decreases as the probability of no-show increases.

The remainder of this chapter is composed of five sections. Section 4.2 examines the literature of appointment scheduling in regards to overbooking and open-access scheduling. Section 4.3 presents the assumptions used to construct the discrete-time, discrete-state, stationary infinite-horizon Markov decision process. Section 4.4 explores the experimental design used to examine the proposed model. Section 4.5 presents the results and analysis of the model. Section 4.6 summarizes and concludes the objectives of this research.

4.2 Literature Review

Appointment systems are designed to minimize waiting times for patients while maximizing the utilization of physicians and other resources. Cayirli and Veral [82] provide an extensive literature review that describes the fundamental factors associated with appointment systems which entails the following: number of services available, number of physicians, number of appointments per clinic session, the arrival process of patients in relation to punctuality, no-shows, walk-ins, service times, lateness and interruption level of doctors, and queue discipline. The authors also present the measures of performance in regards to cost-based, time-based, congestion, fairness, and other appropriate measures. More recently, Gupta and Denton [26] present the challenges and opportunities faced with appointment scheduling. With respect to efficiency and timely access to health services, the authors suggest that appointment systems be designed to balance both the needs of the service provider and the patients. Providers may require specific time slots to be available in the clinic and restrictions on how available slots may be filled. In contrast, patients may have preference in both physician and day/time of week. In addition, Kaandorp and Koole [84] further emphasize the needs of these stakeholders in which physicians prefer to be more productive (less idle time) and patients tend to want shorter waiting time. By using patient classification in the design of appointment systems, Cayirli et al. [82] aim to improve patients' waiting time, physicians' idle time and overtime in the absence of making trade-offs between the patient and provider.

Based on these characteristics, the open-access scheduling system is designed to better match patients with their providers as patients request same-day appointments. Several studies have examined the transition from traditional scheduling to open-access systems that aim to improve efficiency of primary care clinics, reduce no-show rates, manage walk-ins, reduce waiting times for scheduling appointments, and restructure types of appointments and length of appointments [9, 44-55]. Under this paradigm, same-day patients are served within the normal clinical hours. However, healthcare providers are often required to work overtime in the event that all patients are not served within their specified appointment slot. Thus, it is critical that clinics allocate the appropriate percentage of prescheduled and available appointment slots. However, there is still an associated risk in managing no-shows with prescheduled appointment slots and ensuring that too few appointments are available for same-day request. It is worth noting that under the open-access scheduling paradigm the actual percentage of prescheduled and open appointments will vary from clinic to clinic.

Here, this research introduces an approach that is not necessarily based on the patient-physician matchup, but patient-multi-slot preference. This work also examines how overbooking may be used to help mitigate some of the risk that is involved with the acceptance of dependent demand arrivals. As noted by McGill and Van Ryzin [18], overbooking is the oldest and most studied revenue management strategy within the airline industry as a response to controlling the probability of denied boardings. Overbooking is also the most utilized approach of revenue management for patient scheduling in traditional appointment models. Within both industries, there is a general assumption that demand exceeds available supply. With respect to healthcare, this translates to how patient demand exceeds the number of physicians available at a single moment. With overbooking models, the objective is to determine the optimal booking limit for each time period that maximizes expected revenues, as one considers the probability of cancellations and penalties for exceeding capacity [18]. Therefore, we explore how overbooking has been used to both improve clinic efficiency and to mitigate the loss of patients due to no-shows and/or cancellations.

Kim and Giachetti [75] develop a stochastic mathematical overbooking model (SMOM) to determine the optimal number of patient appointments to accept to maximize expected total profits for diverse healthcare environments without incurring overtime cost. SMOM considers the probability distribution of no-shows and walk-ins to obtain an optimal solution for the number of patient appointments to be scheduled. The authors recognize that implementation of a naïve statistical overbooking approach (NSOA), which is based solely on the difference between the average number of no-shows and the number of walk-ins, is easier than SMOM. However, SMOM is proven to be a better and more efficient model, since it requires tracking of patient no-shows, cancellation, and walk-in rates. Their model is limited, since it did not provide advice on how to allocate the extra appointments in the schedule in order to reduce patients waiting times. LaGanga and Lawrence [86] demonstrate how the scheduling complexity increases when appointment overbooking is used to compensate for no-shows. The paper presents a 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. The authors use simulation and regression analysis to show that appointment overbooking in healthcare clinics can have a significantly positive net impact on clinic performance by increasing patient access and improving clinic productivity. This, in turn, translates into reduced clinic costs and improved patient satisfaction and outcomes. In addition, the paper identifies situations in which overbooking is most likely to be beneficial and, conversely, in which it is likely to be counterproductive.

LaGanga and Lawrence [90] extend their earlier work to develop a simulation model to mitigate the loss of productivity of physicians due to patient no-shows by testing the performance of scheduling rules for overbooked appointments. Again using simulation, the authors' primary objective is to analyze the effects of the placement of the extra appointments in an overbooked appointment schedule via double-booking, block scheduling, and wave scheduling policies. They suggest that the challenge with overbooking is determining the appropriate allocation of the extra appointments. The simplest overbooking schedule compresses all inter appointment times by the same show rate factor, which is proven to perform well for the various show rates. To avoid the need to have "catch-up time" or large accumulations of patient wait time, the authors did not recommend scheduling policies with very tight appointment slots at any show rate. The authors determined that patient wait time can be avoided by scheduling one extra appointment at the end of the clinic session when the show rate is 0.9. If less overtime is desired, wave scheduling avoids a large accumulation of patient wait time anywhere in schedule. The authors also suggest that clinics should overbook one extra patient per provider per clinic session.

Overbooking models have also experienced challenges in employee morale and patient satisfaction. Kros et al [104] examine the effects of employee burnout as a response to overbooking patients. The authors suggest that the cost of over-scheduling is directly impacted by the burnout cost imposed on service providers. These healthcare providers are expected to both see more patients and extend their workday with the use of overtime. Burnout also occurs from shorter durations when patients are worked into the schedule. Zeng et al. [92] use a game theoretical approach to model the interactions between healthcare clinics and their patients. Based on the patient's history of no-show, the authors propose a selective dynamic overbooking strategy that is used to determine if the clinic should allow the patient to be overbooked. In addition, the authors implement the naïve statistical overbooking policy that is introduced by Kim and Giachetti [75]. As a result, the authors determined that patients should only be overbooked if the clinics are capable of classifying patients in a manner that can be utilized to segment the patients into different classes based on whether or not overbooking is implementable. More importantly, the authors demonstrate that based on the patient's characteristics overbooking may or may not improve clinic's profit. With overbooking models, it has been noted that overbooking tends to penalize patients that arrive for their scheduled appointments by increasing the amount of time they spend waiting to see their physician [92-93]. Realizing this dilemma, schedulers must identify other approaches that will not negatively impact patient satisfaction.

Given both the benefits and challenges of overbooking, the proposed model seeks to determine the optimal percentage of open slots under predetermined overbooking limits. Unlike the other models, this research explores the use of overbooking within an open-access scheduling paradigm rather than the traditional scheduling system. In addition, the overbooking model is applied at the clinic level rather than the provider level, since double-booking has been considered a norm in traditional scheduling systems. Therefore, the proposed model addresses both the needs of the patients who require multiple appointments and the concerns of the healthcare providers in ensuring that the necessary resources are readily available.

Muthuraman and Lawley [59] also develop an overbooking model under an open-access scheduling system for a single service period to compensate for the probability of no-show for outpatient clinics. However, the authors use multi-objective optimization to develop an overbooking process that minimizes patient wait time, maximizes resource utilization, and minimizes the number of patients waiting at the end of the day. The model is limited by the options available to the scheduler when considering patients' preference in provider. In addition, their model attempts to assign patients consecutively by spacing them well apart to reduce overflow between slots. In fact, the average number of patients that are assigned to the later slots is less than the earlier ones. The authors observed that overtime and waiting costs for additional patients increasingly outweigh additional revenues. The objective function is also maximized when the number of patient types is increased. This can be attributed to the increase in flexibility made available to the decision maker by the large number of patient types. Finally, the model can be used as a prediction tool for clinics seeking to determine their daily profit, since the call-in sequence on the schedule profits exhibit a normal behavior.

This research also aims to model how open-access scheduling systems allow clinics to increase their revenue. Current research in applying revenue management to primary-care clinics has modeled the effect of patient choices in identifying which regular patients to accept or reject in order to serve same-day appointment requests [15]. To the author's knowledge, quantitative models, developed to understand open-access scheduling systems, only explore single independent patient arrivals within predominately single provider scheduling models [6, 15, 57, 59-60, 91, 105]. However, there are a few papers that study the impact of open-access scheduling on multiple provider models [15, 56, 58]. For each model, the authors consider the probability of no-show and the allocation of appointment slots through the use of patient-physician matchup for same-day appointment request. These models also consider patient choice in provider and appointment slot. Gupta and Wang [15] extend their single provider model to identify the optimal booking limit for multiple providers. Using a Monte Carlo simulation method, the model examines several heuristic policies to determine the upper bound of the optimal booking limit, the bounds of the booking limit in the acceptance of physician's appointment slot based on patient class, and the critical numbers for each physician. The model determines that the optimal decision is based on the patient choice and reservation state of the clinic as a whole.

Kopach et al. [58] develop a model to simulate the booking of appointments within an open-access scheduling system for a teaching hospital with multiple physicians.

The simulation model is used to examine the impact of open-access scheduling on the continuity of care and the clinic's throughput. Their results indicate that the fraction of patients using open-access contribute significantly to both performance measures. In addition, the number of patients assigned within the physicians' care group provides significance to the continuity of care. DeLaurentis et al. [56] develop a simulation and queueing model to study the impact open-access scheduling has on the patient's clinical visit. With the queueing model, the authors are interested in the utilization of the physicians and the expected waiting time the patient spends within the clinic. The results of the simulation indicates like Kopach et al. [58] that the percentage of open-access appointment requests is the most significant factor in regards to the continuity of care. Also, the model suggests the number of patients that are able to schedule an appointment with their primary physician declines as the fraction of patient requests increase from zero to 75%. The authors suggest that a primary care team be composed of a single primary and two secondary physicians to maintain the continuity of care under an open-access scheduling system.

Like Liu et al. [91], this research presents an infinite horizon Markov decision process. However, the authors' objective is to determine which day to schedule an individual patient's request to be seen by their physician. Their model considers both the probabilities of no-show and cancellation for clinics where open-access is not implementable and traditional scheduling is current practice. Using simulation, the authors compare their proposed improved open access heuristic to five other scheduling policies. However, the authors do not consider the use of overbooking.

Table 16 demonstrates the contributions that have been made towards quantifying the theory and objectives of open-access scheduling and whether or not revenue management (overbooking) concepts are utilized. More importantly, the table demonstrates how current research in patient scheduling has not considered dependent demand arrivals. Here, the models are studied based on the presence of no-show and/or cancelled appointments, the number of providers, the number of patients seeking to schedule a same-day appointment and dependency of the patients amongst themselves. The key to optimizing appointments is to take a quantitative approach to develop the schedule rather than relying on an expert's experience [6].

Table 16. Quantitative model comparison

Author	*Model*	*Single Provider*	*Multiple Provider*	*Dependent Demand?*	*No-Shows?*	*Appointments*	*Distribution*	*Overbooking?*
Qu [6]	Mathematical	X		No	Yes, known	Total number is known and fixed.	Known for prescheduled and OAS	No
Gupta and Wang [15]	Finite MDP and Heuristics	X	X	No		Considers Patient Choice		Yes
Liu et al. [91]	Infinite MDP and Simulation	X		No	Yes	Proposes Improved OAS for traditional system.		No
Kopach [58]	Simulation		X	No	Yes	Allows double-booking	Poisson	No
Giachetti et al. [57]	Simulation	X		No	Yes			Yes
Muthuraman and Lawley [59]	Multiobjective optimization	X		No	Yes	Allows overbooking. Appointment slot allocation. Patient choice.	Exponential	Yes
DeLaurentis et al. [56]	Simulation and Queueing Model		X	No	Yes			Yes

Based on the literature survey, this research seeks to develop a quantitative model to examine the impact of dependent demand (i.e. batch) arrivals under an open-access scheduling system. To date, quantitative papers in this area consider single and multiple provider models under the assumption that demand arrivals are independent among patients. In regards to revenue management, McGill and Van Ryzin [18] stated that it must consider the inclusion of batch bookings as critical area for research.

4.3 Problem Description

4.3.1 Assumptions

Prior to constructing the Markov decision process model, several assumptions are established in terms of the clinic structure, prescheduled appointments, and batch appointments. In regards to the clinic structure, it is assumed that the capacity is fixed, which represents the number of physicians. There is a single appointment type with a fixed duration. While fixed appointment slots may be restrictive, it is representative of some clinics that perform preventative or routine health services (i.e. dental cleanings and physical examinations that are required for athletes). However, overbooking is used to expand this fixed capacity in which the backlog is constrained to a maximum limit. Here, the maximum backlog serves as an upper bound for the number of appointments that can be overbooked per period. When considering an open-access paradigm, this extra capacity increases the number of available slots; in turn, increasing the likelihood that a batch request will be accepted. The total number of patients in the system is determined by the number of physicians busy and the number of patients waiting in the backlog. In addition, those patients waiting in the backlog mitigates the loss of productivity associated with patients failing to meet their scheduled appointment. It also allows those patients that have been accepted within a batch appointment request to wait until the first available physician is idle. The model does not consider patient-physician matchup. In addition, each physician is homogeneous in workload and service time. Open slots are perishable and cannot be carried forward into the next period.

Under an open-access scheduling paradigm, the probability of the physicians being busy represents the ratio between prescheduled and open appointment slots per period. Here, clinic capacity is analogous with available physicians and thus serves as the maximum number of prescheduled appointment slots in a given time period. The prescheduled appointments have priority over same-day batch appointments. The probability that a prescheduled appointment for physician exists in the next period is defined by a binomial distribution. This assumption is also made by LaGanga and Lawrence [106] to determine the distribution of the number of patients arriving for an appointment slot. In the current period, the probability of no-show/cancellation is known and conditioned on the number of prescheduled appointments. This conditional probability follows a binomial distribution and is based on the probability that the patient will fail to meet their scheduled appointment.

With respect to the batch appointment assumptions, a fixed family size for each request that is a function of the number of available physicians is considered. Each patient within the batch is homogeneous and dependent on one another. That is knowledge that a patient within a batch will not meet their scheduled appointment or cancels, affects the probability that the other patients within the same batch will also not meet their scheduled appointment or cancels. It is worth noting that the current prescheduled appointments do not differentiate between a batch and an individual. As a result, the no-show rates are not dependent on the individuals within the family. A request for a batch appointment that is made in the beginning of the period can be processed in the same period if there is a no-show/cancellation. Thus, accepted patients are immediately available to be served by idle physicians. After the acceptance of a batch request, each patient is seen as an individual and the patient(s) are willing to wait to be seen in the backlog.

The assumptions for the model are largely based on the observed behavior of a pediatric dental clinic in [101]. Clinics that also offer preventative care and routine services for children such as annual flu shots and eye care may also have similar assumptions. In addition, the model accounts for the variation of revenues that is generated through the predetermined reimbursement plan. It is assumed that the clinic absorbs the difference between actual billing price and the reimbursement amount. The operating cost is assumed to be absorbed by the local government. These assumptions are consistent with a clinic, whose primary source of income is from Medicaid.

4.3.2 Mathematical model

A discrete-time, discrete-state, stationary infinite-horizon Markov decision process (MDP) model is developed to examine the impact batch appointment requests have on a clinical scheduling system. Based on the current state of the clinic, the model is used to determine whether to accept a request for a batch appointment for the same period or to reject the request. The optimal scheduling policy and expected profit per period are determined using the policy iterative algorithm developed by Howard [107]. With this model, the scheduling coordinators are able to determine the actual scheduling policy for a given batch appointment request. The MDP is formally defined as follows, where Table 17 summarizes the notations used throughout our formulation.

Table 17. Model notation and descriptions

Notation	Description
C	Capacity of the clinic
M	Maximum overbooking limit
i	Number of physicians busy due to prescheduled appointments
j	Number of patients assigned to the backlog
a	Number of prescheduled appointments assigned to the next period
b	Batch size of appointment request
n	Number of no-show/cancellations
x	Number of appointments required
y	Number of appointments available
$p_{N\|i}(n)$	Probability of no-show of prescheduled appointment given current state i
$p_B(b)$	Probability of arrival of request for batch appointment
$p_A(a)$	Probability of prescheduled appointment assigned to next period
$p(s'\|s)$	Probability of transitioning from current state, s, to future state, s'
s	Current state of clinic for two-dimensional space (i, j)
s'	Future state of clinic for two-dimensional space (i', j')
π_s	Steady-state probability for state s
$r(s,k)$	Expected immediate reward vector
k	Binary value to determine the acceptance of a batch appointment request
δ	Revenue generated from serving a patient
λ_1	Penalty cost associated with carrying a backlog
λ_2	Penalty cost associated with having unutilized appointment slots
$E[S]$	Overall expected number of patients served
$E[PS\|i]$	Expected number of prescheduled patients served
$E[BS\|j]$	Expected number of backlogged patients served
$E[b]$	Expected number of patients backlogged
U	Utilization of physicians

System state: Since this work is interested in the clinic level overbooking limit rather than individual physicians, the state of the system is defined to be the aggregated status of the physicians and the number of patients in the backlog. The reservation state is determined by a two-dimensional state vector *(i, j)*:

$$S = \{(i, j) \mid i = 0,1,..C, j = 0,1,..M\} \tag{4.1}$$

Here, *i* denotes the number of physicians that are busy due to prescheduled appointments at the beginning of the period, *j* denotes the number of patients that are currently assigned to the backlog, and *C* denotes the capacity of the clinic. The number of patients in the backlog ranges from (0…*M*), where *M* is the maximum number of patients that are allowed to wait.

Control alternatives: The set of admissible actions for a given state $s \in S$ is defined by $A_s = \{k, k \in [0,1]\}$. Here, k is an admissible decision in which two alternatives are evaluated: 0- to reject the request for batch appointment and 1- to accept a request for a batch appointment. It is common practice for scheduling coordinators to allow rejected patients to be assigned to the next available appointment slot. Previous models in patient scheduling assume that patients that have preference in both their physician and appointment slot are willing to be scheduled at a later date [15]. This occurs without a penalty being applied to the clinic. Therefore, no penalty is assigned for rejecting patients in the proposed model.

Transition function: Several events have been identified that cause a transition from the current state (i, j) to future state (i', j'). These stochastic events consist of the following: no-show/cancellation of prescheduled appointment $p_{N|i}(n)$, arrival of request for batch appointment $p_B(b)$, and existence of a prescheduled appointment in the next period $p_A(a)$. Given the current state $s = (i, j) \in S$, transition to a future state $s' = (i', j') \in S$ occurs as follows, where $(g \wedge h) = \min(g, h)$:

$$x = j + b * k \tag{4.2}$$

$$y = C - (i - n) \tag{4.3}$$

$$i' = a \tag{4.4}$$

$$j' = (M \wedge [x - y]^+) \tag{4.5}$$

$$p(s' | s) = p_A(a) * p_B(b) * p_{N|i}(n) \tag{4.6}$$

Equation 4.2 determines the number of appointment slots required to serve patients that are currently present in the backlog and that have been accepted. The acceptance of the appointment request is based on the batch size $b \in \{0,2\}$, where zero represents the rejection of patients. Next, equation 4.3 computes the number of actual appointment slots that are available after accounting for the number of no-shows, where n represents the number of no-show/cancellations in the period. Note that y is nonnegative, since no-shows are only permitted when the number of prescheduled patients exceeds zero. Therefore, the number of physicians available can be computed as a function of the number of prescheduled patients that fail to arrive for their scheduled appointment or cancel. Equations 4.4 and 4.5 determine the future state of the system, where i' represent the arrival of prescheduled patients, $a \in \{0..C\}$, into the next period and j' represent the number of patients in the backlog that are waiting to be served in the next period. Equation 4.6 defines the transition probability given the system is in current state s_1 and transitions to future state s_2. The transition equation is a function of the number of prescheduled appointments assigned to the next period, the number of appointment slots required to serve patients, and the number of actual appointment slots that are available. The availability of appointment slots considers the maximum backlogged allowed in the system. Figure 15 displays the timeline of the events associated with the transition from current state $s = (i, j) \in S$ to future state $s' = (i', j') \in S$.

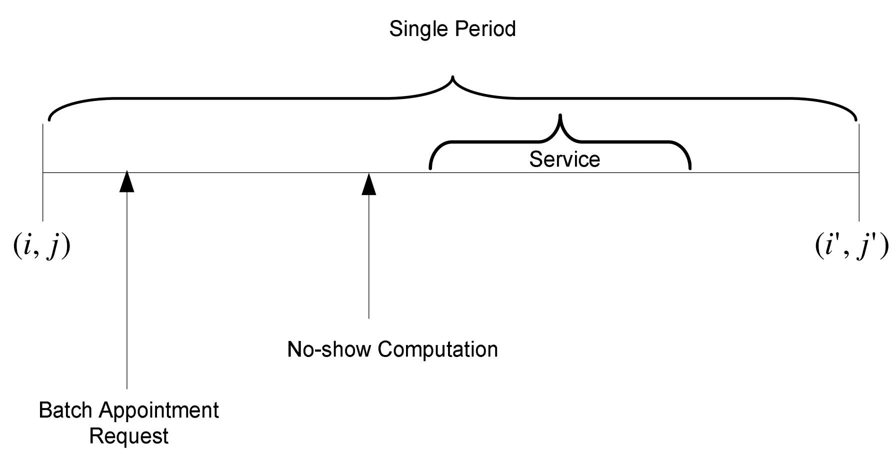


Figure 15. Events timeline

Reward model: The immediate reward associated with current state ($s=(i,j)\in S$) and action (k) is generated as a function of the number of patients being served, the number of unutilized appointment slots, and whether or not there are patients assigned to backlog. The expected immediate reward vector $r(s,k(s))$ is represented by:

$$r(s,k_s)=E[(\delta*((x\wedge y)+(i-n))-\lambda_1*(j+b*k_s-(x\wedge y))-\lambda_2[y-x]^+ \tag{4.7}$$

where the expectation is taken with respect to the number of prescheduled patients (i) and the batch appointment request size (b). The revenue generated from serving a patient is denoted by δ. k_s is the binary decision value indicating the acceptance (1) or rejection (0) of a batch appointment request. Although not modeled directly, the cost of rejection represents both patient dissatisfaction and the inability to serve the daily demand of patients under an open-access scheduling paradigm. Equation 4.7 provides the expected immediate reward with respect to the no-show/cancellation of prescheduled appointment $p_{N|i}(n)$, arrival of request for batch appointment $p_B(b)$, and prescheduled appointment assigned to next period $p_A(a)$.

A stationary optimal control policy is generated, which maximizes the clinic's expected profit per period (g) under an infinite horizon average reward criterion. The optimality equation is expressed in component notation as

$$0 = \max_{k\in A_s}\left\{r(s,k_s) - g + \sum_{s'\in S} p(s'|\, s)v(s') - v(s)\right\}\forall s\in S \tag{4.8}$$

where $v(\cdot)$ can be interpreted as the limiting relative value associated with starting the system in the specific state. In addition, the model also examines the following long-run performance measures: the total expected number of patients that are served, the utilization of physicians, and the expected number of patients in the backlog per period. Please note that we expand *s* to be *(i,j)* to further investigate the model in terms of the number of prescheduled patients and the number of patients that are waiting in the backlog.

As expressed in Equations 4.9 through 4.11, the total expected number of patients that are served is a function of the number of prescheduled and backlogged patients that are served and the optimal steady-state probability π_{ij} associated with a clinic state (i,j). The number of prescheduled served (Eq. 4.10) is computed by the number of patients that fail to show/cancel and the probability of this event occurring. Equation 4.11 represents the expected number of backlogged patients that are served when the batch is accepted. If the request is rejected, only the first term of the equation is computed where the probability of batch arrival is equal to one. Based on the total expected number of patients that are served and the number of physicians at the clinic, Equation 4.12 determines the utilization of the physicians. Equation 4.13 identifies the expected number of patients in the backlog per period. Finally, the optimal policy of the MDP model identifies the ideal scheduling policy for the clinic. With this scheduling policy, the scheduling coordinator will be able to identify whether or not to accept a request for a batch appointment given the prescheduled and open-access appointment rate and the probabilities of no-show and batch arrival. Therefore, the scheduling policy determines the number of patients that are within the system, which is based on the number of prescheduled patients and the number of patients waiting to be seen.

$$E[TS] = \sum_{(i,j)\in S} E[PS \mid i] + E[BS \mid j] * \pi_{ij} \tag{4.9}$$

$$E[PS \mid i] = \sum_{n=0}^{i} (i-n) * p_{N|i}(n) \tag{4.10}$$

$$E[BS \mid j] = \sum_{n=0}^{i} (j \wedge y) * p_{N|i}(n) * p_B(0) + ((j+b) \wedge y) * p_{N|i}(n) * p_B(b) \tag{4.11}$$

$$U = \frac{E[TS]}{C} * 100\% \tag{4.12}$$

$$E[b] = \sum_{(i,j)\in S} j * \pi_{ij} \tag{4.13}$$

4.4 Computational Study

The purpose of this study is to determine the ideal open-access scheduling paradigm that increases the likelihood that a batch appointment request is accepted. Experiments are conducted to examine how the optimal scheduling policy is affected by varying degrees of server idleness, patient behavior, and the appointment request size. In addition, this work explores the effects of overbooking on the performance measures. These experiments provide insight on which ratio of prescheduled and open appointment slots and overbooking limit that best suit the needs of a clinic. Also, this research identifies the relationship between the probability of no-show and probability of batch arrival on the scheduling policy.

Table 18 displays the levels of sensitivity and the associated values for each observed parameter. Although the probability of no-show is varied from zero to one, in practical settings the average no-show rate is 0.20 for families with two children [101].

In addition, the probability of a batch arrival (two or more patients) at the clinic is assumed to be 0.30. Therefore, this sensitivity analysis is performed to examine the impact the change in both the probability of no-show and probability of batch arrival has on the scheduling policy.

Table 18. Parameters for sensitivity analysis

Parameter	Level	Values
Clinic Capacity (*C*)	1	[3]
Maximum Overbooking (*M*)	3	[0, 0.5*C, C]
Maximum Batch Size (*b*)	1	[0.5*C]
Ratio of Prescheduled Appointments ($p_X(x)$)	2	[0.5, 0.7]
Probability of No-show ($p_{N\|i}(n)$)	11	[0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
Probability of Batch Arrival ($p_B(b)$)	11	[0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0]
Revenue (δ)	1	[$135]
Backlog Penalty (λ_1)	1	[$54]
Unutilized Penalty(λ_2)	1	[$13.50]

It is worth noting, when calculating both the maximum overbooking limit and the batch size for the appointment request as half the value of the capacity, the actual value is rounded to the higher integer value. For example, the number of physicians, (*C*), is three, the batch appointment request size, (*b*), is two and the overbooking limit is zero, two and three. With respect to the overbooking limit, zero represents the case in which overbooking is not allowed, while double-booking is implied when the limit is equal to three. The cost parameter is derived based on the actual maximum expected reimbursed revenue provided by the clinic at $135. It is assumed that the backlog penalty is equal to 40% of the revenue; whereas, the penalty for unutilized slots is 10% of the revenue. The model assumes the penalty for unutilized slots will always be less than the penalty for backlogging patients.

4.5 Results

In the following sections, the results of the computational study are presented. For Sections 4.5.1 and 4.5.2, the result are presented under the assumption that the probability of the batch arrival is equal to 0.30, which is in the range of the actual percent of appointment request for batch appointments observed in [101]. The first section characterizes the optimal policies that are generated. The second section summarizes the system's performance measures as the maximum overbooking limit, ratio of prescheduled appointments and probability of no-show are varied. The next section examines the impact of the probabilities of no-show and batch arrival on the optimal scheduling policy. Various capacity levels are explored to determine an optimal overbooking limit, while maintaining a fixed batch size in the final section.

4.5.1 Characterizations of optimal policies

In order to understand the behavior of the system, the structure of the optimal policy is analyzed when the ratio of prescheduled appointments is equal to 0.5 and 0.7. For both scenarios, the policy structure is observed under various no-show probabilities and maximum overbooking limits. The structure of the optimal policy as a function of the no-show rate and the state space is presented in Figure 16. As evident in Figures 16(a) and 16(b), the scheduling policy suggests to reject requests for a batch appointment when the total number of patients in the clinic has exceeded the number of physicians available.

However, it is interesting that the actual rejections occur at different no-show rates. Figure 16(a) experience these rejections when the no-show rate is less than or equal to 0.4; whereas, the same behavior is observed when the no-show rate is less than 0.4 in Figure 16(b). For the actual no-show rates observed at the clinic (0.1 and 0.2), the appointment requests are rejected when the total available capacity (the number of idle physicians and the overbooking limit) is less than the requested batch size. As seen in Figure 16(a), these rejections occur at states (2,3), (3,2) and (3,3). This behavior also occurs when the probability of no-show ranges between zero and 0.1 in Figure 16(b) for states (2,2), (3,1), and (3,2). This implies that the clinic should only reject batch appointment requests when either the overbooking limit has been reached and/or at most one physician is idle. If there is a guarantee that the patients will meet their scheduled appointment (i.e. the no-show rate is 0), then both models tend to reject requests when the total capacity exceeds the number of physicians.

Figure 16(c) illustrates the behavior of the system when overbooking is not allowed (i.e. the maximum overbooking limit is equal to zero). In this case, the model only rejects the appointment request, when the clinic has reached its total capacity and the no-show rate is less than or equal to 0.10. This is important here, because it suggests even in a full system with no predefined overbooking limit (3,0), if no-show rates are greater than or equal to 0.2 a clinic is able to accommodate batch appointment requests. Although batched patients are accepted, the maximum number of patients than can be served is restricted to the number of physicians. In a practical setting, this implies that the physicians will have to work overtime in the event that patients are still waiting in the backlog at the end of the period. This demonstrates the importance of considering the use of overbooking in the initial phases of planning rather than requiring physicians to work overtime unexpectedly. In addition, the clinic must consider how patient satisfaction is impacted, when waiting times are increased due to the acceptance of additional patients.

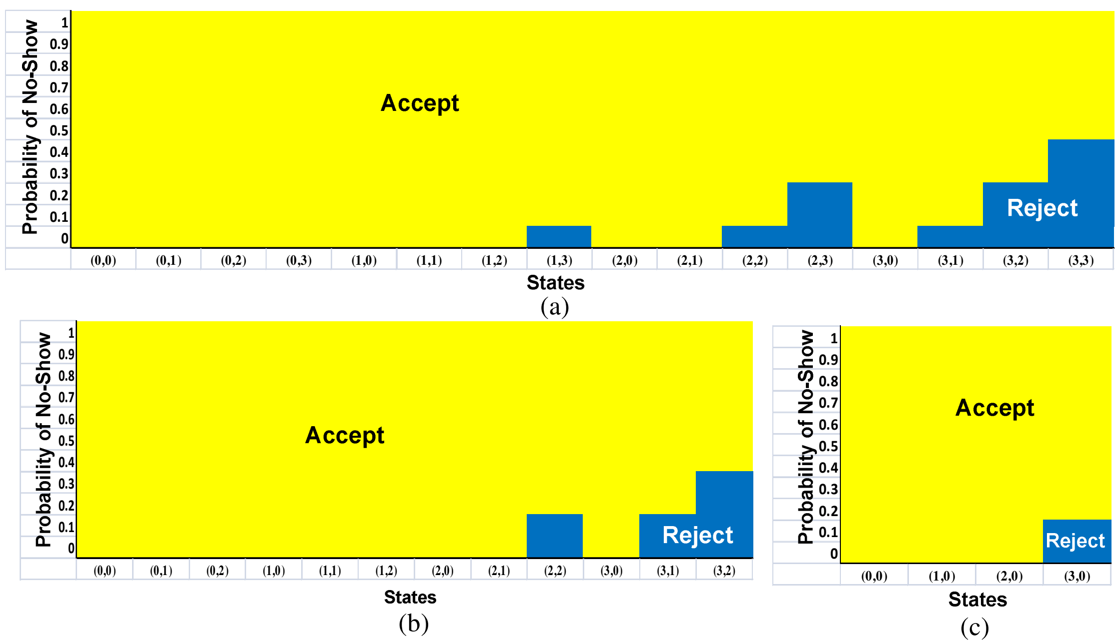


Figure 16. Policies when prescheduled ratio is 0.5 and maximum overbooking limit is three (a), two (b), and zero (c)

Therefore, under an open-access scheduling paradigm that considers the ratio of prescheduled appointments to be 0.5, the likelihood that batch appointments are accepted is high. In fact, the clinic is able to accept patients if only half of the physicians are booked with prescheduled appointments, since the other half is available to serve accepted batched patients. However, rejections do occur when the total number of prescheduled and overbooked patients is greater than the number of physicians. This also implies that acceptance of batch appointment request is not strictly based on the number of patients in the backlog, since patients are always accepted when the backlog has reached its maximum limit and there are no patients prescheduled. Table 19 summarizes the characterizations of the optimal policy in terms of the rejection condition for the respective maximum overbooking limits and the probability of no-show.

Table 19. Rejection region when the ratio of prescheduled appointments is 0.5

M=C		**M=0.5*C**		**M=0**	
$p_{N\|i}(n)$	*Condition*	$p_{N\|i}(n)$	*Condition*	$p_{N\|i}(n)$	*Condition*
0	$i+j>C$	[0, 0.1]	$i+j\geq C+M-1$	[0, 0.1]	$i+j=C+M$
[0.1, 0.2]	$i+j\geq C+M-1$	[0.2, 0.3]	$i+j=C+M$		
[0.3, 0.4]	$i+j=C+M$				

The same analysis is conducted as presented above for the case in which ratio of prescheduled appointments is increased to 0.7. This analysis is important since this is the ratio presented in the literature as the ideal percentage of prescheduled and open appointments. The optimal policies are presented in Figure 17 for the various maximum overbooking limits. As expected, the rejection region increases when the maximum overbooking limit is greater than zero. With Figure 17(a), the model tends to always accept irrespective of the number of patients in the system when the no-show rate is greater than or equal to 0.5. This behavior is observed when the ratio of prescheduled appointments is equal to 0.5 for no-show rates greater than 0.4. Again, rejections for batch appointments occur only when the total capacity is greater than or equal to the number of physicians available on a given day. If there is a guarantee that the patients will meet their scheduled appointment, then each model rejects batch requests when the total capacity is greater than or equal to the number of physicians. This suggests that with the increase in prescheduled appointments, the model is more restrictive. This is based on the fact that under the same condition seen in Figures 17(a) and 17(b), the rejections occurred when total capacity exceeded the number of physicians. However, Figure 17(c) exhibits the same behavior as previously mentioned when overbooking is not allowed.

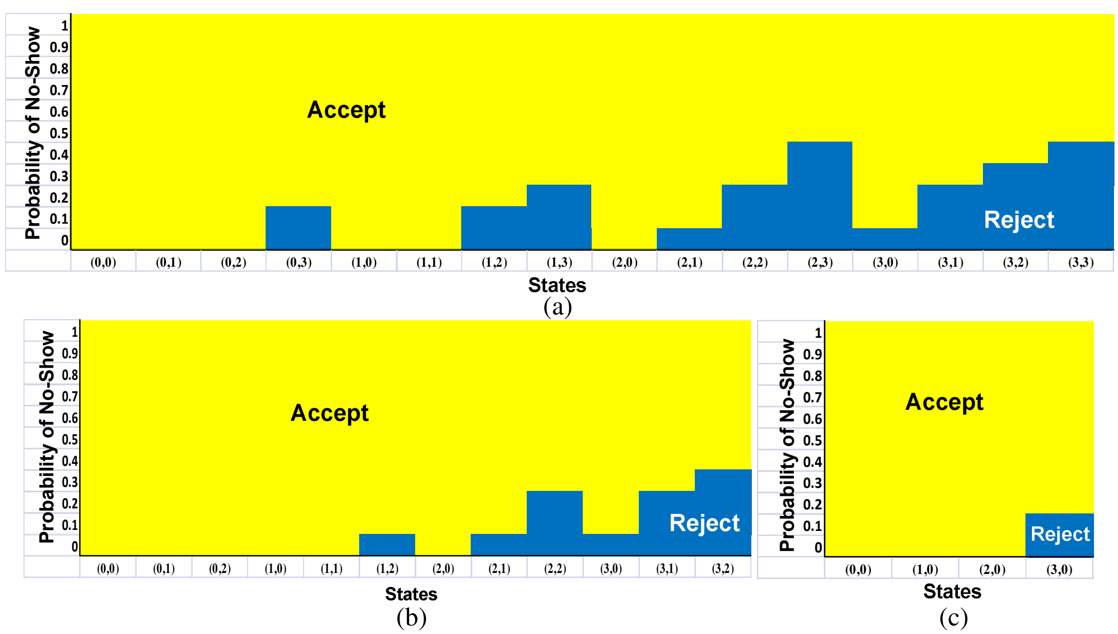


Figure 17. Policies when prescheduled ratio is 0.7 and maximum overbooking limit is three (a), two (b), and zero (c)

Table 20 displays the various characteristics for the rejection region when the ratio of prescheduled appointments is 0.7. Again, the rejection region is based on both the no-show rate and the maximum overbooking limit. From the table, it is evident that rejection region is impacted by the increase in the ratio of prescheduled appointments. In particular, the number of patients that are already assigned to the backlog influences the rejection of batched patients when physicians are double-booked (i.e. the overbooking limit is equal to the number of physicians). In addition, the no-show rate increases the complexity of determining when to reject requests.

Table 20. Rejection region when the ratio of prescheduled appointments is 0.7

M=C		**M=0.5*C**		**M=0**	
$p_{N\|i}(n)$	*Condition*	$p_{N\|i}(n)$	*Condition*	$p_{N\|i}(n)$	*Condition*
0	$i+j \geq C$	0	$i+j \geq C$	[0, 0.1]	$i+j=C+M$
0.1	$\begin{cases} i+j \geq C \,\&\, j \geq M-1 \\ i+j > C \end{cases}$	[0.1, 0.2]	$i+j \geq C+M-1$		
0.2	$i+j > C$	0.3	$i+j=C+M$		
0.3	$i+j \geq C+M-1$				
0.4	$i \geq C-1 \,\&\, j=M$				

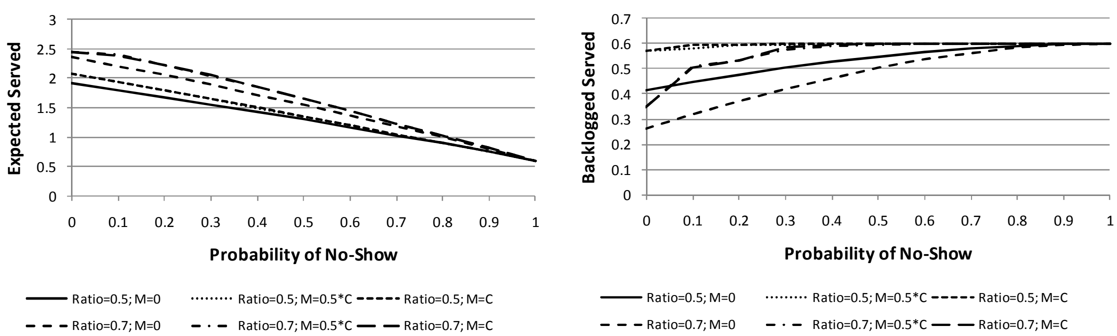
4.5.2 Impact of prescheduled patients and maximum overbooking limit results

Several figures are generated to illustrate the performance of the system. We classify the results based on the ratio of prescheduled and open appointments slots at 0.5 and 0.7 as the maximum overbooking limit is equal to zero, two and three.

4.5.2.1 Expected number of patients served

Figures 18 through 20 do not directly illustrate the behavior of the clinic when the maximum overbooking limit is equal to half of the clinic's capacity. This is a result of the overbooking value being rounded to the higher integer value. Therefore, the system behaves the same when the clinic is overbooked by two and three appointment slots. Based on Figure 18(a), the total expected number of patients that are served decreases as the probability of no-show increases irrespective of the overbooking limit and prescheduled appointments ratio. This is intuitively obvious as the no-show rate decreases the actual demand for resources and thus decreases the output of the physicians (per period). As expected, the clinic is able to serve more patients when the number of prescheduled patients is higher (as measured by the ratio). This implies that the clinic is able to serve patients at higher no-show probability values, since physicians are able to serve those patients that have been assigned to the backlog and/or accepted as a batch appointment request.

Figure 18(b) illustrates the behavior of the system for the expected number of backlogged patients that are served. As the probability of no-show increases, the average number of waiting patients that are served increases but are bounded by the number of physicians available. In addition, waiting patients tend to be served when the physicians are less busy with prescheduled appointments. When the clinic does not allow overbooking, the system is still able to serve patients that have been accepted within the same period. This behavior is observed for both prescheduled appointment ratios and the various no-show probabilities. As expected, the maximum overbooking limit does not have an effect on the average number of prescheduled patients that are served so, we omit this figure. This occurs due to the assumption that the prescheduled patients will be served prior to those patients that are accepted within a batch appointment request. As a result, the expected number of prescheduled patients that are served decreases as the probability of no-show increases for each capacity level.



(a) (b)

Figure 18. Expected number of patients served total (a) and backlogged (b)

4.5.2.2 Expected number of patients assigned to backlog

Figure 19 demonstrates the behavior of the system for the expected number of patients that are assigned to the backlog. As evident in the figure, patients that are assigned to the backlog decreases as the probability of no-show increases. These patients that are waiting represent those patients that will be carried over to the next period as a result of physicians being unable to immediately serve them. This is a result of the system reaching its overbooking limit. Recall in Figure 18(a), the total number of patients that are served is greater at lower probability of no-show values due to the arrival of prescheduled patients and the acceptance of batch appointment request. In general, batch requests are rejected when the total number of patients in the system exceeds the number of physicians available. It is worth noting that the average size of the backlog never exceeds one. The increase in prescheduled appointments leads to a greater increase in patients waiting to be served. The total expected number of patients that are served between 0.0 and 0.1 is 2.5 which is close to the number of physicians available when the capacity is three.

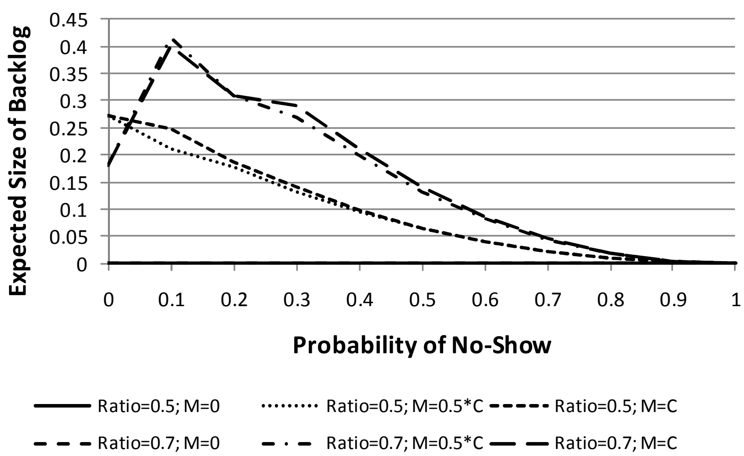
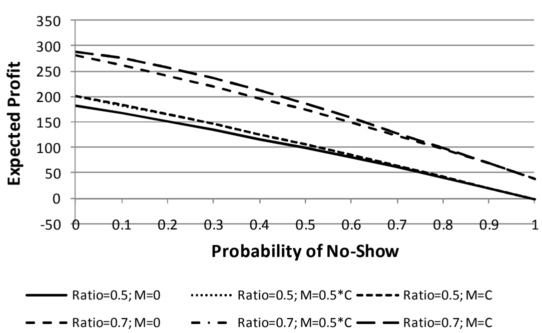


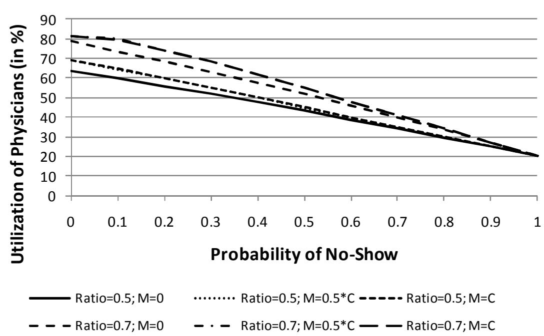
Figure 19. Expected number of patients in the backlog per period

4.5.2.3 Expected profit and physicians' utilization

Figure 20(a) illustrates the expected profit per period as the probability of no-show is increased. By increasing the number of prescheduled patients, an increase in the expected profit per period is realized. As the probability of no-show increases, the expected profit per period decreases monotonically for the respective prescheduled appointments ratios. Regardless of the number of overbooked slots, the clinic continues to remain profitable as the no-show rate is increased, since the batch request is served immediately. However, the clinic is more profitable when overbooking is allowed. This is a direct result of the penalty for unutilized appointment slots being significantly lower than the penalty for patients being backlogged (i.e. left waiting to be served). Therefore, if possible, the clinic should aim to schedule patients in advance versus allowing a higher percentage of open appointment slots for same-day appointment requests. Similar results are displayed in Figure 20(b) with respect to the utilization of physicians. Here, irrespective of the probability of no-show, the higher the ratio for prescheduled appointments; the higher the utilization of the physicians. Thus, the highest utilization (82%) is achieved when the prescheduled appointments ratio is 0.7 and each physician is double-booked within a single period. In practice, it is common for physicians to utilize multiple rooms in the attempt to increase the number of patients that can be served.



(a)



(b)

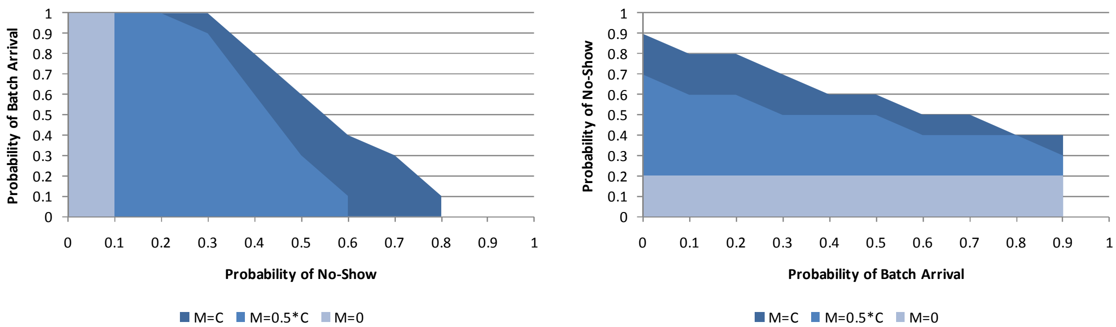
Figure 20. Expected profit per period (a) and utilization of physicians (b)

4.5.3 Impact of batch arrival and no-show probabilities

4.5.3.1 Threshold regions

Here, this research seeks to determine the minimum acceptance and rejection regions as both the probability of no-show and probability of batch arrival are varied when the capacity is three. These regions are generated for a fixed probability value (in the x-direction), while the associated probability along the y-axis is varied. The ratio of prescheduled and open appointment slots is assumed to be equal to 0.5. With respect to the always accept region, a threshold value is identified to provide the minimum value allowed for the respective probability.

Figure 21 illustrates the threshold values for the various overbooking limits for the always accept region for both the probability of no-show and the probability of batch arrival, respectively. This threshold region increases as the maximum overbooking limit increases. In addition, the always accept threshold region tends to decrease as the probability of no-show increases. In fact, this figure can assist scheduling coordinators in determining which batch requests to accept given the probabilities of no-show and batch arrival appointment requests. For example, if the clinic has no prior knowledge of the patterns of the batch arrivals, then a request will always be accepted when the probability of no-show is less than or equal to 0.3. However, given the clinic has knowledge of the probability of batch arrivals, then the scheduling coordinator can determine which patients to accept based on their no-show rate and the clinic's overbooking limit.



(a) (b)

Figure 21. Always accept threshold

In general, rejections occur when the number of patients in the system (both prescheduled and backlogged) is greater than or equal to the clinic's physician capacity level for both the probability of no-show and probability of batch arrival analysis. Similar to the always accept region, the minimum rejection region provides the minimum probability for rejecting batch requests for the various total capacity (TC) values, which includes both the clinic capacity and maximum size of the backlog. As evident in Figure 22, the minimum rejection region increases as the total capacity increases as well as the respective probability value. Like the always accept threshold, the scheduling coordinator can determine which batch request to reject given information on both probabilities or only one and the total capacity of the clinic. Although not illustrated, the scheduling coordinator will reject requests when the probability of a batch arrival is less than or equal to 0.9 and the probability of no-show is less than or equal to 0.1. This behavior occurs when the overbooking limit is equal to zero (i.e. the total capacity is three).

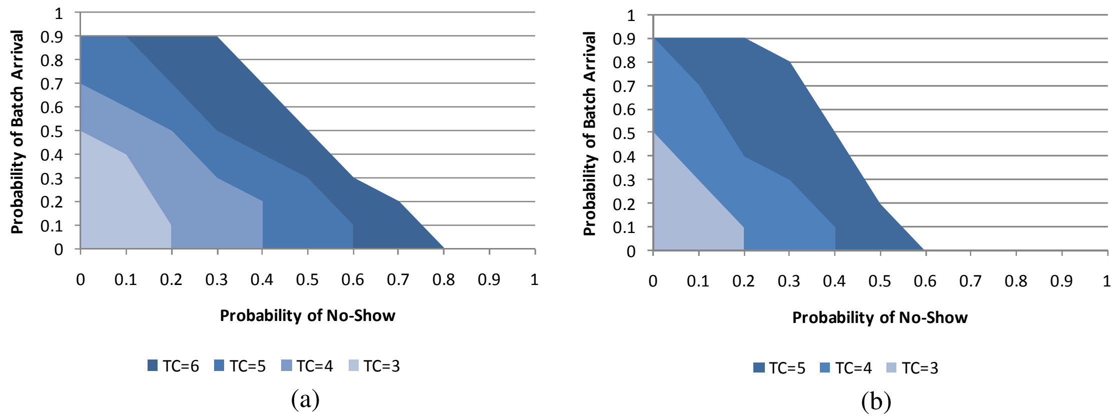


Figure 22. Minimum rejection threshold when the overbooking limit is equal to three (a) and two (b)

4.5.3.2 Effects on performance measures

To further investigate the impact of the no-show and batch arrival probabilities, the behavior of the system is studied under the predetermined performance measures. For a fixed probability of no-show, there is a similarity in the system's performance for the following measures: the expected profit per period, the physicians' utilization, total expected number of patients that are served, and the expected number of patients that are served from the backlog. As the probability of batch arrival increases, each of these measures tends to also increase. In fact, the backlog tends to be affected by the variation of the probability of batch arrival. This is based on the assumption that the accepted patients are willing to wait to be served by the first available physician. Therefore, the expected number of patients that are assigned to the backlog increases as the probability of batch arrivals increases. In addition, the expected number of prescheduled patients that are served remains constant, since it is only affected by the probability of no-show.

In contrast, each of the performance measures, except for the expected number of patients that are served from the backlog, decreases as the probability of no-show increases and the probability of batch arrival remains fixed. This is a direct result of the idle physicians being able to serve the waiting patients immediately. These results suggest that the variation of probability of batch arrival positively impacts the utilization of the clinic, when the probability of no-show is constant. Our results also indicate that the variation in probability of no-show has a negative effect on the utilization of the physicians. In general, the backlog tends to be the most affected by the variation of both the probabilities of no-show and batch arrival. This is due to the clinic accepting requests for batch appointments for higher no-show rates. In addition, the always accept region is slightly higher for a fixed probability of no-show versus a fixed probability of batch arrival. Similar results are also identified for the minimum rejection region.

4.5.4 Exploration of optimal overbooking limit

The purpose of this experiment is to determine the optimal overbooking limit for various capacity levels at the clinic. Under a perfect scenario (i.e. each physician is busy with a prescheduled patient and each patient meets their scheduled appointment), the clinic should overbook only the batch size value. Hence, the maximum overbooking limit is equal to the size of the batch appointment request. However, in the experiments presented thus far, the overbooking limit is assumed to be fixed. In a practical setting, each physician would not be required to have patients overbooked in each appointment slot. For a fixed batch size of two and by varying the clinic's capacity from two to four, the optimal overbooking limit is identified for each of the respective capacity values. The optimal overbooking limit is determined by identifying the overbooking value that maximizes the expected profit per period for each capacity level. Again, the ratio for prescheduled appointments is assumed to be equal to 0.5 and the probability of batch arrival is equal to 0.3.

Table 21 provides the optimal overbooking limit for the respective values. In most cases, the optimal overbooking limit surpassed the desired overbooking limit as the no-show rate is increased. In fact, Figure 23 displays the difference between observed optimal overbooking limit and the batch size value. Here, the optimal overbooking limit is closer to the desired value at higher no-show rates. This implies that the number of physicians that have to serve additional patients declines when prescheduled patients fail to meet their scheduled appointment. It is worth nothing that the difference is less than 0.17 for each capacity value except for the case when the capacity is equal to 4 and the probability of no-show is 0 (at 0.58).

Table 21. Optimal overbooking limit given capacity and probability of no-show

	Probability of No-Show				
Capacity	*0*	*0.2*	*0.4*	*0.6*	*0.8*
2	C-1	C	C+1	C+1	C+2
3	C-1	C	C+1	C+2	C
4	C-1	C+1	C+1	C	C-2

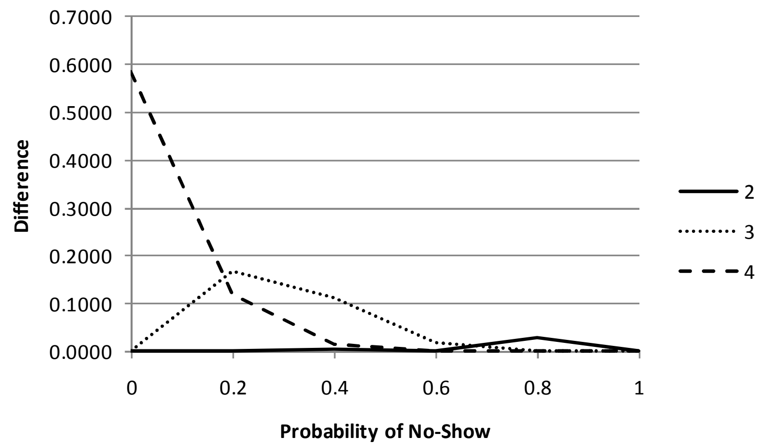


Figure 23. Difference between desired and optimal overbooking limit

This research also examines the scheduling policy for the respective capacity, probability of no-show and optimal overbooking limit values. Table 22 provides the condition in which batched patients are accepted and rejected. Similar to the results presented in Section 4.5.2, appointments are rejected when the total number of patients in the system is greater than or equal to the clinic's capacity. Under the capacity used for the base model (C=3), the condition remains the same when the probability of no-show ranges from 0 to 0.6. Thereafter, the optimal scheduling policy always accepts the request for batch appointments. A capacity of three is the only value that exhibits a linear relationship as the probability of no-show increases. With the other two values, the highest optimal backlog is achieved when the probability of no-show is 0.6 (C=3) and 0.4 (C=4).

Table 22. Scheduling policy for optimal overbooking limit

	C=2			C=3			C=4		
p_{Nli} *(n)*	*M**	*Decision*	*Condition*	*M**	*Decision*	*Condition*	*M**	*Decision*	*Condition*
0.0	1	Reject	$i+j\geq C$	2	Reject	$i+j\geq C+M-1$	3	Reject	$i+j\geq C+M-1$
0.2	2	Reject	$i+j\geq C+M-1$	3	Reject	$i+j\geq C+M-1$	5	Reject	$i+j\geq C+M-1$
0.4	3	Reject	$i+j\geq C+1 \& j\geq C$	4	Reject	$i+j\geq C+M-1$	5	Reject	$i+j=C+M$
0.6	3	Reject	$j=M$	5	Reject	$i+j\geq C+M-1$	4	Always accept	
0.8	4	Reject	$j=M$	3	Always accept		2	Always accept	

4.6 Conclusion

This chapter has presented a framework to illustrate dependent patient demand arrivals under a multiple provider model. A discrete-time, discrete-space, stationary infinite-horizon Markov decision process (MDP) model has been developed to study the impact that batch arrivals have on an open-access scheduling system. Current research in open-access scheduling considers single provider models, where demand arrivals are independent among patients. This research is motivated by a pediatric dental clinic who must determine whether to accept a request for a batch appointment for a same day slot. The optimal scheduling rule obtained from the MDP model suggests the batch appointment requests should always be accepted regardless of the number of patients prescheduled and the maximum overbooking limit, when the probability of no-show/cancellation for each prescheduled appointment is greater than or equal to 0.50.

The clinic should always reject a request for batch appointment when the probability of no-show is less than or equal to 0.10 and the number of patients either with a physician and/or waiting in the backlog is greater than the number of physicians in the clinic. Additional rejections of appointment request also occur as the probability of no-show increases when either the backlog has reached its maximum limit and/or each physician is busy with a patient that meets their prescheduled appointment. However, in spite of the probability of the no-show, the scheduling policy suggests to accept request when the number of patients either with a physician and/or waiting in the backlog is less than the number of physicians in the clinic.

Like Kros et al. [104], this work determined that overbooking can be used to increase patient access to care. In addition, the results indicate that the clinic should only consider using overbooking in their scheduling policy if the probability of no-show is low and the probability that the physicians are busy is high. Similar results have been noted by LaGanga and Lawrence [90], as they identified that clinics should overbook one extra appointment at the end of the clinic session per provider when the probability of no-show is low. In general, if the clinic continues to reserve half of their appointment slots for same-day requests, then the system is capable of serving the request for batch appointments within the same period regardless of the number of overbooked slots. However, clinics, who currently schedule appointments under an open-access paradigm, should allocate 70% of their appointment slots for prescheduled patients. At this percentage, the clinic will be more profitable as well as experience an increase in the utilization of physicians and the number of patients that are served. With an increase in the ratio of prescheduled appointments, there is an increase in the number of batch appointments that will be rejected. In addition, clinic should not consider overbooking patients when the probability of patients not showing up is low. Thus, by allowing the patients to wait in the backlog, there is only a slight impact in increasing the total expected number of patients that are served, the utilization of the physicians, and the expected number of patients waiting to be served.

Finally, clinics must rely on information pertaining to the probability of no-show, the probability of batch arrival request, and ratio of prescheduled appointments to determine whether or not a request should be accepted. In addition, the clinic must determine the optimal number of patients that they are willing to allow to wait to be seen by an idle physician in the backlog and the actual amount of the penalty associated with the patient waiting. This sentiment is also addressed by LaGanga and Lawrence [86], as their results suggest that clinics should understand the impact of both patients no-show behavior and their cost structure prior to making general statements about overbooking. This work has an underlying limitation in that model was not validated at the pediatric dental clinic motivated by research due to unforeseen changes. Future research areas include determining the actual optimal scheduling rule based on various batch sizes and examining the allocation of patients to physicians within the scheduling paradigm. Recall, this research assumed that the batch size was half of the number of physicians in the clinic. This work also considered the physicians as an aggregated unit, in which each physician is homogenous. The model can also be extended to identify the impact of batch arrival requests in more dynamic scheduling techniques that is not limited to an open-access scheduling paradigm. With the change in scheduling approaches, healthcare providers can gain further insight of the impact of dependent demand arrivals.

CHAPTER 5

Finite-Horizon Stochastic Model to Determine the Effects of Scheduling Independent versus Dependent Patients

5.1 Introduction

Traditionally, appointment systems are designed under the assumption that patients are independent in terms of their arrivals and no-show rates. However, this assumption is often invalid when families seek to schedule multiple appointments within the same time frame. The concept of multiple appointments in patient scheduling is discussed in [101], in the context of pediatric dental clinics. The term batch is introduced to account for multiple patients requesting service from a set of providers with the primary expectation that all patients will be processed simultaneously and/or consecutively. In addition, these patients are dependent amongst each other in terms of both their arrival and no-show probability. Due to the complexity of scheduling multiple patients within a single period, scheduling coordinators must ensure that resources are available to meet the demands of the patients seeking care. This concept is not unique within the healthcare industry field. Dependent demand arrivals also exist within the hotel and airline industries, as well as in product manufacturing. However, in the healthcare setting, this is a relatively unexplored topic.

Although, batch appointments can increase the productivity and utilization of the clinic, there are several challenges that may arise with the acceptance of batch appointments. First, if batched patients are late for their appointment, then the clinic may experience longer waiting times for other patients. In addition, given that the batch fails to arrive without calling (i.e. no-show), then the number of unutilized slots will be significantly higher than that encountered with individual appointment requests.

To negate the possibility of physicians being idle due to batch patients not showing up, we propose the use of overbooking. First introduced by the airline industry in the 1960s, overbooking models determine the optimal booking limit for each time period that maximizes the expected revenue, while considering both the probability of no-show/cancellations and penalties for exceeding capacity [18]. Overbooking models are designed to accept more reservations than the available capacity level [75]. LaGanga and Lawrence [86] suggest that overbooking be used as a means to mitigate the negative impact of no-shows, improve patient access, and increase provider productivity for healthcare clinics. In addition, overbooking is used to stabilize the revenue streams for clinics by seeing more patients within a period [92].

Although, the healthcare and airline industries have similar objectives in their use of overbooking, they differ significantly in their approaches to handling overbooked patients. Unlike the healthcare industry, the airline industry has a fixed capacity in which they must consider fair alternative arrangements for overbooked patients [104]. However, the healthcare industry is able to expand their capacity with the use of overtime. In addition, patient no-shows occur throughout the day; whereas, no-shows for the airline customers all occur at a single point in time [86]. Moreover, the healthcare industry has very different cost structures than the airline industry in which the primary source of payment is through reimbursement from patients' health maintenance organization (HMOs), government programs like Medicaid or Medicare, or out-of-pocket expenses.

A major challenge with overbooking models is finding ways to balance the risk associated with too few patients showing up and too many patients showing up [75]. Overbooking may also contribute to prolonged patient waiting, which in turn, may negatively impact patient satisfaction [92]. Muthuraman and Lawly [59] suggest that the longer wait time is a direct result of more patients arriving to be seen, which causes the clinic to experience excessive workloads. As a result, the clinic is faced with substantial changes to its systems dynamics due to an increase in the number of patients that are overflowed from slot to slot throughout the day. Thus, implementers of overbooking must acknowledge that this technique is more appropriate when the product or service being sold is perishable for a fixed capacity, which is difficult or too expensive to change in a short term [75].

The objective of this research is to study the acceptance of individual and batch appointment requests during a finite-horizon using stochastic dynamic programming. The model is used to determine if batch appointments negatively impact the clinic's performance (i.e. expected profit and number of patients that are served. In addition, this work examines if overbooking is necessary to increase the likelihood of batch appointments being scheduled. Lastly, this research identifies how the performance measures are impacted by the acceptance criteria: individual patients only, batch appointments only, and a hybrid of both patient demand types.

The remainder of this chapter is organized as follows. Section 5.2 examines the literature of appointment scheduling systems. Section 5.3 presents the assumptions used to construct the finite-horizon stochastic dynamic program. Section 5.4 explores the experimental design used to examine the proposed model. Section 5.5 presents the results and analysis of the model. Section 5.6 summarizes and concludes the objectives of this research.

5.2 Literature Review

Patient scheduling has been extensively studied by Cayiril and Veral [82], who present the fundamental factors of outpatient scheduling based on literature from the 1950s to early 2000s. More recently, Gupta and Denton [26] present some of the ongoing challenges and potential research areas for appointment systems. Therefore, readers are directed to their work as this research focuses on overbooking models. Only one model to date has examined the impact of batch appointments on patient scheduling systems [95]. However, a greater emphasis is placed on how overbooking models are implemented under the assumption that patients are independent in their arrival and no-show rates. Research in healthcare overbooking is composed of simulation, analytical, and mathematical models. These overbooking models are characterized based on their objective, the use of overtime, the allocation of overbooked patients, the number of overbooked slots, and the results or insights from the respective model.

Using simulation, LaGanga and Lawrence [86] examine the use of appointment overbooking to reduce the negative impact of no-shows. The authors also seek to develop a method in which clinical decision makers can determine if they should implement the use of overbooking. Thus, the authors construct an analytical utility model that evaluates overbooking in terms of trade-offs between the benefits of serving additional patients and costs of increased wait time and provider overtime. The model assumes the clinical session can be overbooked beyond its normal capacity in which the time interval between sessions is reduced proportionately to the number of sessions overbooked. The authors acknowledge that overbooking may yield an increase in patient waiting times and an increase in clinic overtime even when considering patient no-shows. However, these shortcomings are disproportionately greater for small rather than larger clinics. The authors also recognize that when using overbooking, provider productivity declines as no-show rates increase, but at a rate that is much lower than without overbooking. Their model also proves that with large no-show rates, overbooking can provide robust productivity performance results.

LaGanga and Lawrence [90] extend their earlier work to explore various methods in which overbooking can be implemented via double-booking, block scheduling, and wave scheduling policies. Again using simulation, the authors' primary objective is to analyze the effects of the placement of the extra appointments in an overbooked appointment schedule. They suggest that the challenge with overbooking is determining the appropriate allocation of the extra appointments. In particular, they examine the following strategies: adjusting time intervals between appointments, using block scheduling of multiple patients at one or more scheduled times, and a combination of those approaches to fit extra appointments into the schedule. For a given show rate, the best schedule is determined by the scheduling rule in itself and the relative performance of the alternative approaches. The simplest overbooking schedule compresses all inter appointment times by the same show rate factor. This method is proven to perform well for the various show rates. To avoid the need to have “catch-up time” or large accumulations of patient wait time, the authors did not recommend scheduling policies with very tight appointment slots at any show rate. In addition, the block scheduling of multiple patients at the same time are not recommended unless the patients show rate is less than or equal to 0.5. The model also determined that patient wait time can be avoided by scheduling one extra appointment at the end of the clinic session when the show rate is 0.9. If less overtime is desired, wave scheduling avoids a large accumulation of patient wait time anywhere in schedule. The authors also suggest that clinics should overbook one extra patient per provider per clinic session.

Kros et al. [104] present an analytical study to determine the potential costs and benefits of overbooking for a clinic appointment schedule. Their model aims to investigate the use of overbooking along with the perceptions and acceptance of its use from the healthcare providers. Unlike traditional models that assume costs of overbooking are constant over time per overscheduled patient, the authors consider costs as a nonlinear function of the overbooking rate and employee burnout. The employee burnout is a result of healthcare providers having to work overtime to ensure that all patients are served. In addition, the authors predict the number of patients scheduled and the proportion of no-shows. By modifying the model of LaGanga and Lawrence [86], the authors develop an overbooking model that consists of both the clinic’s scheduling algorithm and the burnout model. Based on the results of the proposed model, overbooking recommendations are implemented in a clinic. The authors suggest that the clinic double-book one morning and one afternoon appointment for each provider working a full schedule, as needed. The goal is to increase the overbooking rate as the clinic gains confidence and skill. Overall, the clinic viewed overbooking as a means to increase healthcare access without having to increase staffing levels and facility space.

Kim and Giachetti [75] propose a stochastic mathematical overbooking model (SMOM) to determine the optimal number of patients to accept to maximize the expected total profit. Unlike the other models presented thus far, the authors' goal is to increase the number of patients that can be seen without incurring overtime cost. In addition, their model considers both probability distributions for no-shows and walk-ins rather than just the probability of no-show. The authors recognize that implementation of a naïve statistical overbooking approach (NSOA), which is based solely on the difference between the average number of no-shows and the number of walk-ins, is easier than SMOM. However, SMOM is proven to be a better and more efficient model, since it requires tracking of patient no-shows, cancellation, and walk-in rates. Unlike some of the other models, the authors did not provide advice on how to allocate the extra appointments in the schedule in order to reduce patients waiting times.

Chakraborty et al. [108] use multi-objective optimization to examine a sequential clinical scheduling model for patients with a general service time distribution and multiple no-show probabilities. They also consider the use of overbooking to ensure that all patients will be seen in the event of patients not being serviced within the specified appointment slot. In addition, the model aims to balance the reward and costs for patient waiting and staff overtime. Their proposed model obtains a higher expected profit than the seven appointment rules proposed by Cayirili and Veral [82]. Similarly, Muthuraman and Lawley [59] also use multi-objective optimization to develop an overbooking process that minimizes patient wait time, maximizes resource utilization, and minimizes the number of patients waiting at the end of the day. With the use of overtime, the patients that are waiting at the end of the day are served. The model is limited by the options available to the scheduler when considering patients' preference in provider. In addition, their model attempts to assign patients consecutively by spacing them well apart to reduce overflow between slots. In fact, the average number of patients that are assigned to the later slots is less than the earlier ones. The authors observed that overtime and waiting costs for additional patients increasingly outweigh additional revenues. The objective function is also maximized when the number of patient types is increased. This can be attributed to the increase in flexibility made available to the decision maker by the large number of patient types. Finally, the model can be used as a prediction tool for clinics seeking to determine their daily profit, since the call-in sequence on the schedule profits exhibit a normal behavior.

Zeng et al. [92] use a noncooperative game theory model to understand the interactions between healthcare clinics and their patients. Based on the patient's history of no-show, the authors propose a selective dynamic overbooking strategy that is used to determine if the clinic should allow the patient to be overbooked. In addition, the authors implement the naïve statistical overbooking policy that is introduced by Kim and

Giachetti [75]. As a result, the authors determined that patients should only be overbooked if the clinics are capable of classifying patients in a manner that can be utilized to segment the patients into different classes based on whether or not overbooking is implementable. The authors also suggest that clinics monitor and evaluate patients continuously based on their no-show record (i.e. identify habitual no-show patients versus patients that tend to always meet their scheduled appointment).

Based on the literature survey, overbooking has been proven to be beneficial to scheduling coordinators who aim to increase the productivity of their physicians when patient no-shows are present. However, overbooking models must balance the benefits of serving additional patients and the costs associated with patient waiting and physician overtime. Prior models in patient scheduling consider patients to be independent in both their request for appointments and their no-show rate. These models are predominately single provider models. Therefore, a finite-horizon stochastic dynamic program is presented, which considers both patient no-shows and physician overtime in a multiple provider model. This research aims to understand the effects of appointment demand type (individual versus batch) and overbooking on the proposed patient scheduling model. Unlike the models discussed in this section, patients are classified based on their appointment demand type and not their no-show probability. However, each of the respective demand types has their own probability of no-show. In addition, the use of overbooking is considered at every appointment slot, given the influx of demand due to families requiring multiple appointments.

5.3 Problem Description

A finite-horizon stochastic dynamic program (SDP) is presented to examine the acceptance of various patient demand types. The model divides the scheduling system into stages (i.e. appointment slots/sessions) in which a decision must be chosen at each stage. Based on the decision, the clinic is able to understand how the state at the current stage transforms into the state at the next stage [109]. That is identifying the number of patients that are prescheduled and have been accepted as a means of overbooking. The overall objective of the model is to maximize the clinic's expected profit.

The model is constructed under several assumptions. For the clinic structure, there is a fixed capacity, where each physician is overbooked up to one slot per physician per period. However, the number of patients that are served cannot exceed the number of physicians available at the clinic. In addition, the number of patients that are prescheduled is constrained by the clinic's capacity. The backlog represents the overbooked patients that are allowed to wait until the first available physician is idle. Therefore, those patients waiting in the backlog mitigates the loss of productivity associated with patients failing to meet their scheduled appointment. It also provides the clinic with the extra capacity needed to accept additional patients. The model does not consider patient-physician matchup. Each physician is homogenous in workload and service time. The model assumes a single appointment type in which the duration of the appointment and the number of periods are fixed.

This research considers two patient demand types: individual and batch. Each appointment demand type has its own probability of no-show. Patients, who are prescheduled as a batch, are assumed to be dependent. Therefore, if one patient breaks the entire batch breaks. Prescheduled appointments are known in advance (prior to the beginning of the period). Patients that fail to meet there scheduled appointment are realized at the beginning of the period. The number of no-shows is based solely on the number of prescheduled patients for each patient class, which follows a binomial distribution.

At the beginning of each period, the clinic must determine how many patients to overbook. These patients are determined based on the acceptance of either an individual or batch appointment request. The model assumes that the accepted requests arrive at the beginning of the *k*th period and are scheduled to be served in next period, $k+1$. At the terminal period, there is no decision being made. However, patients that are not served in their initial appointment slot will be served at the end of the planning horizon via overtime at a higher cost. As a result, physicians are required to work overtime. Figure 24 illustrates the timeline of these events for a single period.

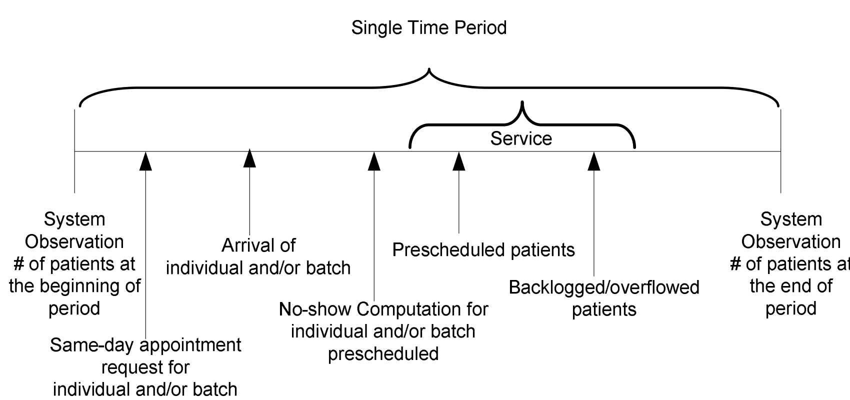


Figure 24. Events timeline

Using a modified version of Bertsekas [110] notation, the SDP is formally defined as follows. Let $\boldsymbol{x_k}$ denote the number of patients in the clinic at the beginning of the *k*th period. This is represented by the aggregated status of the physicians in terms of the number of prescheduled individual patients (z_{1I}) and batch patients (z_{1B}), and the number of individual (z_{2I}) and batch (z_{2B}) patients in the backlog. For the SDP model developed in this paper, the reservation state is determined by a four-dimensional state vector $(z_{1I}, z_{1B}, z_{2I}, z_{2B})$:

$$x_k = \{(z_{1I}, z_{1B}, z_{2I}, z_{2B}) \mid z_{1I} = 0,1..C, z_{1B} = \{0,b\}, z_{2I} = 0,1,..M, z_{2B} = 0,1,..M\} \quad (5.1)$$

$$\text{Where } z_{1I} + z_{1B} \leq C, z_{2I} + z_{2B} \leq M$$

Here, C denotes the capacity of the clinic and M is the maximum number of patients that are allowed to wait. Therefore, the number of patients assigned to the backlog is also constrained. The maximum backlog serves as an upper bound for the number of appointments that can be overbooked per period. The number of patients scheduled for a batch appointment is based on whether or not a batch request is accepted. Therefore, the value of z_{1B} is restricted to either zero or the batch size (b). However, the number of patients that are overbooked as a batch, z_{2B}, can vary from zero to M. This is due to the possibility of the entire batch not being able to be served in the current period. Therefore, the clinic assumes that prescheduled patients have priority over overbooked patients. In addition, patients that are overbooked are willing to wait until the first available physician is idle. In the event that both individual and batched patients are waiting, the individual patient has priority over the batched patients. This ensures that families remain together until everyone is serviced.

The decision, $\boldsymbol{u_k}$,, is defined as the number of accepted patients during the kth period, where the decision is constrained to $u_k \leq M - (z_{2I} + z_{2B})$. The model evaluates four alternatives: (0,0) to reject any request for an appointment, (1,0) to accept a request for an individual appointment, (0,2) to accept a request for a batch appointment and (1,2) to accept both demand types. It is worth noting that in clinical settings, the rejected patients are scheduled for a future date and time. This work only assumes that the rejection of appointment request signify that the clinic is unable to serve the overbooked patients within the next period. Although u_k is represented by two-dimensions, the decision is expanded to $u_k = \{(0,0,d) \mid d = (0,0),(1,0),(0,2),(1,2)\}$. This expansion is necessary to ensure that both the number of individual (z_{2I}) and batched (z_{2B}) patients that are waiting and/or accepted are included. The backlog is updated based on the decision.

Two events occur that cause the transition from the current state, $\boldsymbol{x_k}$, to the future state, $\boldsymbol{x_{k+1}}$: patient arrivals and no-show. The patient arrivals are deterministic for each patient class. Therefore, the presence of no-shows is the only stochastic event that occurs, in which the random variable $\boldsymbol{w_k}$ is defined as the number of patients that do not show during the kth period. Let n_{1I} and n_{1B} represent the possible no-show values for each of the respective demand types. The model assumes that the probability of no-show differs for individual $p(w_k \mid z_{1I})$ and batch appointments $p(w_k \mid z_{1B})$. As previously mentioned, the probability of no-show for individual patients is based on the binomial distribution (Eq. 5.3). However, the probability distribution for batched patients is strictly based on the Bernoulli trial in which the group of patients either shows ($1 - p(w_k \mid z_{1B})$) or does not show ($p(w_k \mid z_{1B})$). If no patients are prescheduled, then the probability of no-show is equal to 1. The effects of each of these events on the respective demand types are expressed as follows.

$$w_k = \{(n_{1I}, n_{1B}, 0, 0) \mid n_{1I} = 0,1..z_{1I}, n_{1B} = \{0, z_{1B}\}\} \tag{5.2}$$

For individual patients:

$$P(W_k = n_{1I}) = \binom{z_{1I}}{n_{1I}} p(w_k \mid z_{1I})^{n_{1I}} (1 - p(w_k \mid z_{1I}))^{z_{1I} - n_{1I}} \tag{5.3}$$

For batched patients:

$$P(W_k = n_{1B}) = p(w_k \mid z_{1B})^{n_{1B}} (1 - p(w_k \mid z_{1B}))^{b - n_{1B}} \text{ for } n_{1B} \in \{0, b\} \tag{5.4}$$

Thus, the future state is function of the number of patients in the clinic, x_k, the number of patients that do not show, w_k, and the number of accepted patients, u_k. The future state and transition probability are derived as follows.

$$y_k = x_k - w_k = ((z_{1I} - n_{1I}), (z_{1B} - n_{1B}), z_{2I}, z_{2B}) \tag{5.5}$$

$$o_k = \begin{cases} (0,0), & if \ \left|y_k\right|_1 \le C \\ (\left|y^*_k\right|_1 - C, z_{2B}), & if \left|y^*_k\right|_1 > C \\ (0, \left|y_k\right|_1 - C), & if \ \left|y_k\right|_1 > C \end{cases} \tag{5.6}$$

Where $\left|y^*_k\right|_1 = \sum_{n=1}^{3} \left|y_n\right|$

$$x_{k+1} = (z_{1I}^{k+1}, z_{1B}^{k+1}, (o_k + u_k)) \tag{5.7}$$

$$p(x_{k+1} \mid x_k) = p(w_k \mid z_{1I}) * p(w_k \mid z_{1B}) \tag{5.8}$$

Equation 5.5 determines the number of patients in the system that require care after the patients that fail to meet their appointment are accounted for. As noted, y_k is reduced to a scalar in which the L^1-norm is computed as $|y_k|_1 = \sum_{n=1}^{4} |y_n|$. If the total number of patients exceeds the number of physicians, then the patients that are unable to get served is expressed in Equation 5.6. Recall, the individual backlogged patients have priority over the batch backlogged patients. Therefore, the backlog for individuals is updated by the total number of prescheduled patients (post no-show computation) and the existing individual backlog. If every patient within the individual backlog are served, then the batch backlog is updated. As evident in Equation 5.7, the future state, x_{k+1}, is based on both the prescheduled appointments for the future period and the number of patients that have been carried from the previous period. Those patients that have been carried are a function of the number of patients that are unable to be served in the current period, o_k, and the acceptance of overbooked patients, u_k. In addition, o_k is equal to zero if $|y_k|_1$ is less than or equal to the number of physicians. Equation 5.8 computes the transition probability given that the system is in current state x_k and will transition to future state x_{k+1}.

The reward functions are formulated as follows, where $(c \wedge d) = \min(c,d)$:

$$J_N(x_N) = g_N(x_N) = (r - \lambda_2) * (z_{2I} + z_{2B}) \tag{5.9}$$

$$J_k(x_k) = \max_{u_k \in U_k(x_k) w_k} E\{g_k(x_k, u_k, w_k) + J_{k+1}(f_k(x_k, u_k, w_k))\}, \quad k = 0,1,...N-1 \tag{5.10}$$

$$g_k(x_k, u_k, w_k) = r * (C \wedge |y_k|_1) - \lambda_1 * (C - |y_k|_1)^+ - \lambda_2 * (|y_k|_1 - C)^+ \tag{5.11}$$

$$Total \quad Accepted = \sum_{k=0}^{N-1} | u_k |_1 \tag{5.12}$$

Where k is the discrete time index and N is the planning horizon. With dynamic programming models, the optimal expected profit is calculated using backward recursion in which the model computes the reward starting at period $N-1$ to the ending period at 0. Equation 5.9 represents the terminal reward incurred at the end of the clinical workday which encompasses both the revenue, r, obtained from serving the remaining backlog and the penalty cost for the physicians' having to work overtime, λ_2. Equation 5.10 calculates the expected reward with respect to the probability distribution of w_k given the clinic has x_k patients in the system at the start of period k. This expectation is a function of the state of the system (x_k) and the decision (u_k). In addition, Equation 5.11 generates the profit that is accumulated over time, k, based on the total number of patients in the system that requires care. Equation 5.11 is also composed of the following cost parameters: the revenue generated from serving a patient (r), the cost associated with having unutilized appointment slots (λ_1), and the cost associated with the physicians' overtime (λ_2). For each period, the model also examines the following performance measures: the total expected profit incurred, $J_0(x_0)$, and the number of patients that are accepted/overbooked. The latter is determined by the total number of patients that are accepted in each period over the entire planning horizon (Eq. 5.12).

5.4 Computational Study

To explore the proposed model, several experiments are conducted to examine the acceptance of individual versus batch appointment requests. The initial state of the clinic is known and the clinic operates from 8:00 AM to 5:00 PM. Therefore, the planning horizon is equal to eight when the appointments are scheduled for a duration of one hour. Since a portion of the state space is represented by the prescheduled appointment type (which is needed to determine the no-show rates), several sample prescheduled states are generated over the planning horizon. Table 23 provides the actual prescheduling sequence for each period and experiment, where periods 0 to 3 represent the morning session and periods 4 to 7 the afternoon.

For the first six schedules, this work considers each possible prescheduled state and assumes it is the same for every period. Schedules 7 and 8 represent the cases where each physician is fully utilized for half of the planning horizon; whereas, 7 considers prescheduled individual patients and 8 considers a prescheduled individual and batch appointment. Schedules 9 and 10 consider only one demand type for half of the planning horizon as well, but the clinic is not fully utilized (to capacity). Next, schedules 11 and 12 represent the cases in which one demand type is scheduled for half of the planning horizon and the other demand type is scheduled thereafter. In the final analysis, different combinations in which each physician is scheduled with either only individual patients or both demand types are represented by schedules 13 through 15.

For each scheduling sequence, the optimal scheduling policy is identified based on the acceptance of individual patients only, batch appointments only, and a hybrid of both patient demand types for various probabilities of no-show. With both the individual and batch only models, the clinic is constrained to only accepting the respective decision or rejecting the request entirely. However, the model still considers each demand type for the prescheduled patients. These acceptance criterions are also examined for the best, observed and worst case scenarios, which are based on the no-show probabilities for each of the patient demand types. The best case represents the scenario in which every prescheduled patient meets their prescheduled appointment (i.e. the probability of no-show is equal to zero); whereas, the worst case assumes every patient fails to meet their prescheduled appointment (i.e. the probability of no-show is equal to one). The observed case is based on the respective probabilities of no-show for batch and individual patients at a public dental clinic, where individual patients experienced a no-show rate of 0.187 and patients scheduled in a batch of two at 0.208 [101].

Table 23. Prescheduled appointments sequence when planning horizon is eight

Schedule	Period 0 8:00 AM	Period 1 9:00 AM	Period 2 10:00 AM	Period 3 11:00 AM	Period 4 1:00 PM	Period 5 2:00 PM	Period 6 3:00 PM	Period 7 4:00 PM
1	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
2	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)	(0,2)
3	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)	(1,0)
4	(1,2)	(1,2)	(1,2)	(1,2)	(1,2)	(1,2)	(1,2)	(1,2)
5	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)	(2,0)
6	(3,0)	(3,0)	(3,0)	(3,0)	(3,0)	(3,0)	(3,0)	(3,0)
7	(3,0)	(3,0)	(3,0)	(3,0)	(0,0)	(0,0)	(0,0)	(0,0)
8	(1,2)	(1,2)	(1,2)	(1,2)	(0,0)	(0,0)	(0,0)	(0,0)
9	(0,2)	(0,2)	(0,2)	(0,2)	(0,0)	(0,0)	(0,0)	(0,0)
10	(1,0)	(1,0)	(1,0)	(1,0)	(0,0)	(0,0)	(0,0)	(0,0)
11	(1,0)	(1,0)	(1,0)	(1,0)	(0,2)	(0,2)	(0,2)	(0,2)
12	(0,2)	(0,2)	(0,2)	(0,2)	(1,0)	(1,0)	(1,0)	(1,0)
13	(3,0)	(3,0)	(3,0)	(3,0)	(1,2)	(1,2)	(1,2)	(1,2)
14	(3,0)	(1,2)	(3,0)	(1,2)	(3,0)	(1,2)	(3,0)	(1,2)
15	(1,2)	(3,0)	(1,2)	(3,0)	(1,2)	(3,0)	(1,2)	(3,0)

In addition, this research determines the impact of the acceptance of individual only, batch only, and both demand types on the total expected profit incurred and the number of patients that are accepted. These experiments provide insight on the acceptance of each of the respective patient demand types that will best suit the needs of a clinic. The optimal scheduling policy is also generated for each possible state. The cost parameters are derived based on the actual maximum expected reimbursed revenue provided by the clinic at $135. It is assumed that the overtime penalty is equal to 40% of the revenue; whereas, the penalty for unutilized slots is 10% of the revenue. The model assumes the penalty for unutilized slots will always be less than the penalty for physicians' overtime. Table 24 summarizes the levels of sensitivity and the associated values for each observed parameter.

Table 24. Parameters for sensitivity analysis

Parameter	Level	Values
Clinic's Capacity (*C*)	1	[3]
Maximum Backlog (*M*)	1	[C]
Maximum Batch Size (*b*)	1	[2]
Individual Probability of No-show $p(w_k \mid z_{1I})$	3	[0, 0.2, 1.0]
Batch Probability of No-show $p(w_k \mid z_{1B})$	3	[0, 0.2, 1.0]
Revenue (*r*)	1	[$135]
Overtime Penalty (λ_1)	1	[$54]
Unutilized Penalty(λ_2)	1	[$13.50]

5.5 Results

5.5.1 Scheduling policies

Identifying the optimal number of overbooked patients is essential to understanding the behavior of the system. Tables 25 through 27 display the structure of the optimal policy for each of the scheduling sequences and probabilities of no-show. This research examines these policies for the best, observed, and worst case scenarios. In practice, the scheduling coordinator may decide to overbook the initial appointment slot to ensure that the physicians are not idle for multiple periods (if patients do not show up for the earliest appointment slot). In this case, the total number of patients that are overbooked can vary from zero to three. Therefore, this analysis is geared to the scenario in which every physician is double booked in initial period.

5.5.1.1 Best case

In regards to the best case scenarios, Table 25 demonstrates how the acceptance of each demand type varies for each schedule and period. As evident from the table, there is a time lag in which patients are overbooked. This is due to the fact that the accepted patients are not served until the next period and that the decision to overbook patients is based primarily on the number of patients in the existing backlog. This behavior is more evident with schedules 1, 2, 5, and 7 through 11. Schedule 1 is the only schedule that accepts three patients (both individual and batch requests) at every other period. Here, each physician is idle for each period, which in turn, guarantees that the overbooked patients will be served in the next period. This scheduling sequence is idle for cases in which the scheduling coordinator is only accepting same-day appointment requests. Schedules 7 through 10 also experience similar behavior when there are no prescheduled patients from periods 4 through 7. Again, this sequence is idle for clinics who restrict when prescheduled patients are scheduled. Thus, the clinic should only accept both demand types when there are enough physicians to serve everyone.

Table 25. Optimal scheduling policies for best case

Schedule\Period	0	1	2	3	4	5	6	7
1	(1,2)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
2	(1,0)	(0,0)	(1,0)	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)
3	(0,0)	(0,2)	(0,0)	(0,2)	(0,0)	(1,0)	(1,0)	(0,0)
4	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
5	(1,0)	(0,0)	(1,0)	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)
6	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
7	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
8	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
9	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
10	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
11	(1,0)	(0,0)	(1,0)	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)
12	(0,0)	(0,0)	(0,2)	(0,0)	(0,0)	(1,0)	(1,0)	(0,0)
13	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
14	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
15	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
Legend:		(0,0)		(1,0)		(0,2)		(1,2)

Schedules 2, 5, and 11 represent the cases in which only individual patients are accepted. This can attributed to the fact that each of these scheduling sequences has two patients prescheduled as either individuals or as a batch from periods 4 to 7. This suggests that the acceptance of overbooked patients is primarily influenced by the scheduling sequence of patients in the afternoon. In addition, the number of patients that are carried in the backlog from period to period is reduced as the time in day increases. Both schedules 3 and 12 accept batched patients in the morning; whereas, individual patients are booked in the afternoon. These schedules both have only one patient scheduled in the afternoon. However, schedule 12 has a batched group in the morning; whereas schedule 3 has only a single patient. Under these scheduling sequences, patients are overbooked in consecutive periods (5 and 6). This implies that the model is aiming to reduce or avoid the long-term penalty associated with carrying a large number of patients in the backlog at the end of the day. In regards to the other schedules, the model rejects all appointment requests when each physician is scheduled with either all individual patients or a combination of both individual and batched patients.

5.5.1.2 Observed case

At a glance, we notice that the best and observed case scenarios presented very similar scheduling policies, which is evident from Table 26. For both cases, schedule 1 accepts three patients (both individual and batch requests) at every other period; while schedules 7 through 10 also accept three patients but in the afternoon sessions only (periods 4 and 6). Schedules 3 and 12 overbook batched patients in the morning; whereas, individual patients are booked in the afternoon. In addition, only individual patient request are accepted in periods 0, 2, and 4 for schedules 2, 5 and 11.

However, due to patients failing to meet their scheduled appointment, the model only overbooks an additional patient for schedule 5 at period 5. Although the probability of no-show in a natural setting seems slightly high, the model does not take on the risk of overbooking patients when the total number of patients exceeds the number of physicians. The total number of patients is a function of both the number of prescheduled patients and the number of patients waiting in the backlog. As a result, the number of patients that are left waiting at the end of the appointment slot is expected to decline as the probabilities of no-show for each demand type increases.

Table 26. Optimal scheduling policies for observed probabilities

Schedule\Period	0	1	2	3	4	5	6	7
1	(1,2)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
2	(1,0)	(0,0)	(1,0)	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)
3	(0,0)	(0,2)	(0,0)	(0,2)	(0,0)	(1,0)	(1,0)	(0,0)
4	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
5	(1,0)	(0,0)	(1,0)	(0,0)	(1,0)	(1,0)	(0,0)	(0,0)
6	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
7	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
8	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
9	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
10	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
11	(1,0)	(0,0)	(1,0)	(0,0)	(1,0)	(0,0)	(0,0)	(0,0)
12	(0,0)	(0,0)	(0,2)	(0,0)	(0,0)	(1,0)	(1,0)	(0,0)
13	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
14	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
15	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)

Legend: (0,0) (1,0) (0,2) (1,2)

5.5.1.3 Worst case

The final case represents the event in which the prescheduled patients are guaranteed to not show up for their scheduled appointment. From Table 27, the model only accepts patients when no one is scheduled as a prescheduled appointment. This is illustrated by schedules 1 and 7 through 10. These schedules also experienced the same behavior in the other two cases. In fact, it seems counterintuitive that the model never overbooks patients for the remaining schedules, given the high no-show rate for prescheduled patients. However, by examining the expected reward for each period, the model is not profitable and never recovers from the patients initially failing to meet their scheduled appointments at the end of day (given the model is computed from the end of the day until time 0). Thus, this implies that by examining the optimal policies is only the initial phase of understanding the behavior of the model.

Table 27. Optimal scheduling policies for worst case

Schedule\Period	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**
1	(1,2)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
2	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
3	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
4	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
5	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
6	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
7	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
8	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
9	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
10	(0,0)	(0,0)	(0,0)	(0,0)	(1,2)	(0,0)	(1,2)	(0,0)
11	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
12	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
13	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
14	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
15	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)	(0,0)
Legend:		(0,0)		(1,0)		(0,2)		(1,2)

5.5.2 Financial impact

To investigate the impact of the no-show probabilities, the behavior of the system is studied in terms of the expected profit. Figure 25 illustrates the total expected profit for each schedule sequence and probability of no-show. As expected, only schedule 1 proves to generate the same expected profit regardless of the no-show rate. Again, this is due to the fact that there are no patients prescheduled during this scheduling sequence. This scheduling sequence is ideal in clinical settings that only allow patients to make request on the same-day that they require services.

The figure also suggests that clinic is most profitable when the prescheduled patients are guaranteed to be present at their appointment. However, only schedules 14 and 15 yield the smallest expected profit under this condition. These schedules represent a combination of all individual patients or either both an individual and one batched group, where each physician is scheduled with a patient at every period. Under these scheduling sequences, the model is still receiving compensation from the number of patients that are waiting at the terminal period. This notion generates an interesting perspective of the impact of the overbooking patients in advance in order to mitigate any loss in patients failing to meet their scheduled appointment (in this case 3). For the remaining schedules, there is a decline in the expected profit as the probabilities of no-show are increased. This is expected given that the clinic is profitable when prescheduled patients meet their scheduled appointment and additional patients are served when overbooking is allowed.

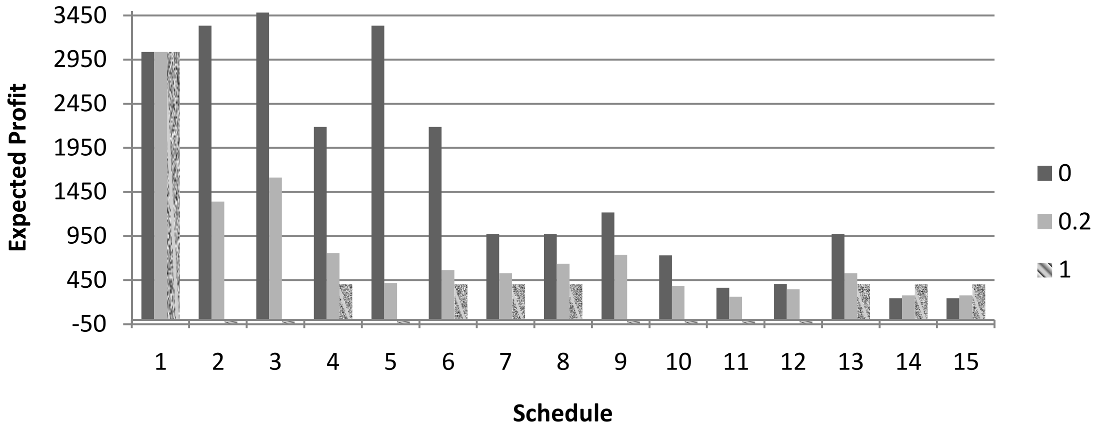


Figure 25. Expected profit when the probabilities of no-shows are varied

5.5.3 Effects of appointment demand type

In the previous section, the effects of the appointment demand type on the proposed model are examined. In this section, the system is explored in terms of the acceptance criteria under the predetermined performance measures. Figure 26 displays the total number of patients that are accepted during the planning horizon for each scheduling sequence. Figure 26(a) suggests that the hybrid model (which considers every possible decision) does not necessarily generate the largest number of overbooked patients than the models that are restricted to accepting individual only or batch only. In fact, the batch only model accepted the most patients for schedules 2, 3, 10, and 11. With each of these schedules, only one appointment demand type was scheduled in a single period. In addition, the model tends to accept the additional batched patients when the total number of patients in the system is less than the number of physicians or when the number of waiting patients is less than the batch size. Again, this is not surprising given the additional resources that are needed to serve both patients. However, by accepting batch appointments as overbooked rather than prescheduled patients, helps mitigate the possibility of these patients failing to meet their scheduled appointments. This, in turn, increases the productivity of the physicians but it decreases the total expected profit since the physicians may have to work overtime to ensure every patient is served.

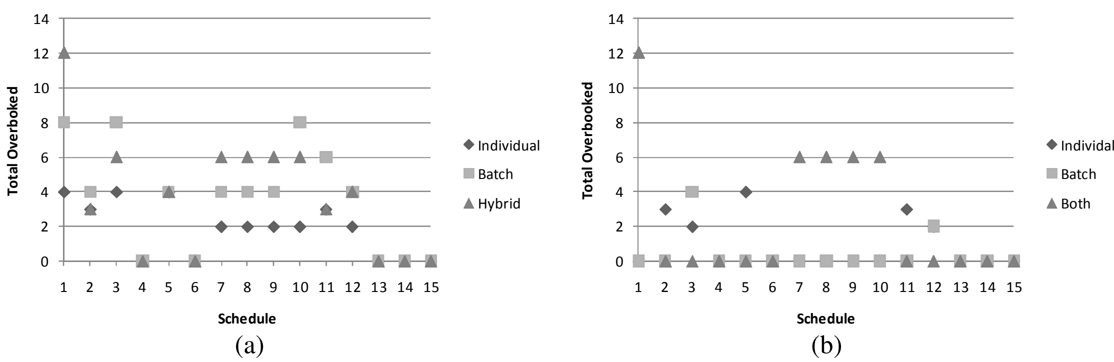


Figure 26. Total number of overbooked patients during planning horizon

From the figure, the individual only case never accepts more patients than the other two models. Since individual patients are independent among each other, the acceptance of a single appointment request is less likely if there is at most one physician idle. However, the models performed the same for schedules 4 through 6 and 13 through 14. Schedule 5 accepted four patients regardless of the model type; whereas, the other schedules did not overbook anyone.

As with Figure 26(b), the acceptance for the hybrid model is decomposed by whether or not a single demand type is accepted or both. More individual patients are accepted for schedules 2, 5, and 11, where the majority of the prescheduled appointments are for two individuals or a batched group in the afternoon (periods 4 to 7). Schedules 3 and 12 accepted more batched patients where only one individual is prescheduled from periods 4 to 7. This suggests that the model is highly influenced by the scheduling sequence of patients in the latter part of the day (i.e. afternoon). Both demand types were only accepted in schedules that did not have prescheduled patients (schedules 1 and 7 through 10). In general, patients are not overbooked when the total number of patients in the clinic exceeds the number of physicians and/or when the backlog has reached its capacity. This is also attributed to the increase in overflowed patients which reduces the objective function significantly. Therefore, the clinic has to balance the risk associated with patients not showing up rather than incurring the additional cost of physician overtime by accepting overbooked patients.

Based on each of the acceptance criteria, Figure 27 compares the expected profit for the respective schedules. In general, there is only a slight difference among the acceptance criteria for most of the schedules. Similar to figure 3(a), the individual only model produced the least in terms of the expected profit performance measure. The hybrid model is the highest with schedules 1, 3, and 9; whereas, the batch only model is the highest with schedules 2, 5, 7 and 8. These results coincide with those found in figure 26(a) in terms of the total number of patients that are overbooked. Therefore, the hybrid model performs better than the models that restrict the acceptance criteria when additional patients could have been overbooked due to multiple physicians being idle. This also suggests that clinic can benefit from overbooking batched patients if the risk is too immense to schedule this demand type in advance. However, scheduling coordinators must minimize the amount of time overbooked patients have to wait to be seen.

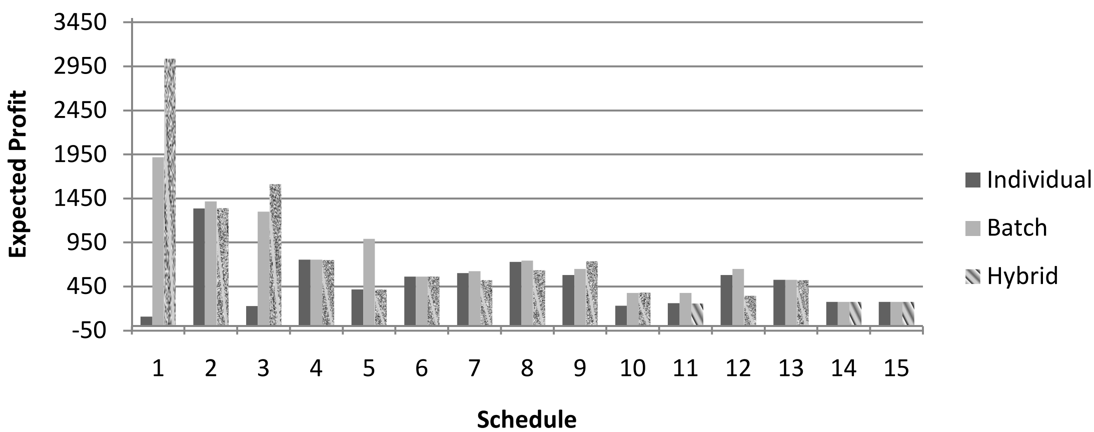


Figure 27. Comparison of expected profit when decision is restricted

5.5.4 Variation of no-show rates for demand types

Thus far, the underlying assumption is that the no-show rates for individual and batched patients are equivalent. However, it is unlikely that the clinic will always experience the same behavior from individual and batched patients. Therefore, this section examines the impact of various probabilities on the total expected profit and the total number of patients that are accepted when the prescheduled patients exhibit the sequence for schedule 12. Here, batched patients are scheduled in the morning session;

whereas, a single individual patient is booked in the afternoon. This scheduling sequence provides a better understanding of how each demand type influences the acceptance of overbooked patients. Figures 28 and 29 represent the results of varying the respective probability of no-show when the accompanying probability of no-show is equal to 0.2 (the observed value).

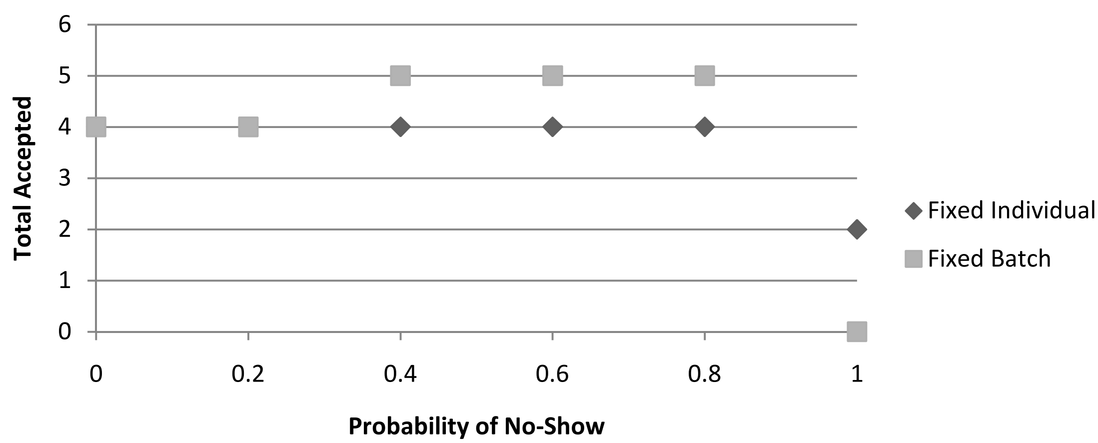


Figure 28. Overbooked patients when probabilities are varied

With Figure 28, the total number of accepted patients is the same for each demand type when the no-show rate is low. Here, four patients are accepted when the probability of no-show ranges from 0 to 0.2. In addition, both demand types accepted a batch appointment request in the morning (period 2) and individual appointment requests in the afternoon (periods 5 and 6). This implies that when a batch is prescheduled, no-one is accepted until there is a chance that everyone is served within the next couple of periods.

In this case, the entire batch that is accepted (in period 2) is served in period 3, if the prescheduled batched patients fail to meet their scheduled appointment. Otherwise, one patient within the batch is served in period 3, while the other patient is not served until period 4. This suggests that the waiting time for batch patients is minimized.

The figure also illustrates how more patients are accepted as the no-show rate for individual patients is varied from 0.4 to 0.6, while the no-show rate for batched patients remains fixed. This is expected given that the general assumption with batch patients is that if one person breaks their appointment than the entire batch will break. In fact, the optimal policy suggests that patients only be accepted in the afternoon, where only one patient is prescheduled. As the no-show rate for individuals is increased, batch appointment requests are accepted in periods 4 and 6; whereas, only one patient is scheduled in period 5. This is implies that more patients are accepted when there is a chance that all of the patients will be served in their allocated time slot. Again, this is achieved with the acceptance of both individual and batched patients throughout the planning horizon. However, the total number of patients that are overbooked is consistent until the no-show rate is equal to 1 when the no-show rate for individual patients is fixed. This is due to the physicians being idle at each period which negatively impacts both their expected profit and utilization.

As evident in Figure 29, the total expected profit differs significantly for batched versus individual patients as the respective probability of no-show is varied. For a fixed no-show rate for individual patients, the figure illustrates how the expected profit decreases as the no-show rate for batched patients increases. Again, this is a result of the effects of batched patients failing to meet their scheduled appointment. When this occurs, the physicians are less productive which in turn negatively impacts the clinics efficiency. However, the clinic is able to recover some of the loss of profit with the use of overbooking as seen in Figure 28. This demonstrates how critical it is for clinics to understand their patient's no-show behavior prior to accepting and scheduling batch appointments.

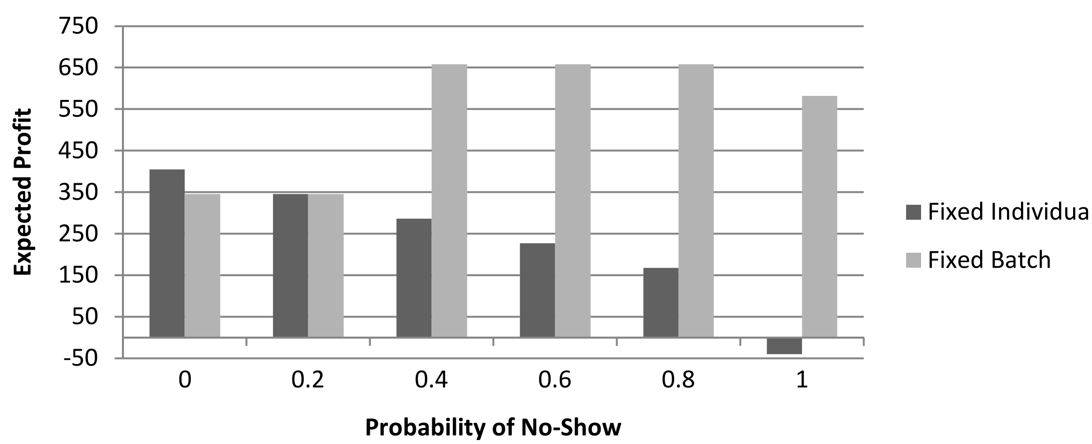


Figure 29. Expected profit when probabilities are varied

However, the clinic can remain profitable in the event that the no-show rates for individual patients are varied and the no-show rate for batched patients is fixed. More importantly, the variability in batched patients has a greater influence on the clinic's profitability than the variability in individual patients. Based on these results, the clinic can remain profitable given the variation of probability of no-show for each of the appointment demand types when overbooking is allowed. However, in terms of the clinic's performance measures, the impact of the variability of individual patients yields better results than that that observed by batched patients. This is attributed to fact that batched patients are dependent among each other, which in turn, increases the complexity of scheduling these patients. Thus, when considering multiple appointment demand types, the scheduling coordinator must optimize their ability to schedule each of the demand types in a manner that increases both the clinic's profitability and physicians' productivity.

5.6 Conclusion

This paper has examined the effects of scheduling independent versus dependent patients. A finite-horizon stochastic dynamic programming model has been developed to study the impact of these appointment demand types on the clinic's profitability and the physicians' productivity. Current research in patient scheduling does not consider the influences of dependent demand. Therefore, this research demonstrated the use of overbooking to mitigate the risk associated with patients being dependent among each other in terms of their arrival and probability of no-show. In general, when the scheduling coordinator is certain that prescheduled patients will fail to meet their appointment, then the acceptance of overbooked patients is strictly based on the number of patients that are currently waiting in the backlog. However, if prescheduled patients are known to show up as planned, then the acceptance of overbooked patients is based on the total number of patients in the system (i.e. prescheduled and waiting). In fact, the optimal policy determined from the model suggests that rejections only occur when the total number of patients in the system is greater than or equal to the number of physicians in the clinic.

Due to the uncertainty of patients meeting their scheduled appointment, batched patients are overbooked when the total number of patients in the system is less than the number of physicians or when the number of waiting patients is less than the batch size. This is due to the influx in demand which requires additional resources to be needed in order to ensure that the entire group will be served. However, by accepting batch appointments as overbooked rather than prescheduled patients helps, mitigate the possibility of these patients failing to meet their scheduled appointments. This, in turn, increases the productivity of the physicians but it decreases the total expected profit since the physicians may have to work overtime to ensure every patient is served.

In addition, the clinic can remain profitable given the variation of probability of no-show for each of the appointment demand types when overbooking is allowed. However, in terms of the clinic's performance measures, the impact of the variability of individual patients yields better results than that observed by batched patients. Therefore, when considering multiple appointment demand types, the scheduling coordinator must optimize their ability to schedule each of the demand types in a manner that increases both the clinic's profitability and physicians' productivity.

Finally, this work has demonstrated both the need to consider independent and dependent patients and the benefits of overbooking. Future research areas include determining the actual optimal scheduling rule based on various batch sizes and examining the allocation of patients to physicians within the scheduling paradigm. Recall, we assumed that the maximum number of overbooked appointments are known. As a result, the model is limited to various batch sizes. Therefore, additional research in this area can consider the best approach to balance the needs of batched patients and the healthcare providers. The model can also be extended to explore the effects of when only one appointment demand type can be overbooked in each period. With the change in scheduling approaches, healthcare providers can gain further insight of the impact of dependent demand arrivals.

CHAPTER 6

Conclusion

The intent of this research was to introduce and explore the concept of batch appointment scheduling. First, this work determined the prevalence of batch appointments at public and private pediatric dental clinics, in which, patients that are scheduled within a batch are proven to be dependent among each other. This dependency affects both patients' arrival and no-show rate. In fact, the clinics studied for this research experience batch appointment requests at nearly 42%. This research also determined that overall patients at the private clinic had a significantly lower no-show rate at 2%; whereas the public clinic experienced a no-show rate of 18%.

Using multivariate statistical analysis, this work identified (for both clinics) if there is a relationship among the appointment demand type, patient behavior, and reason for the appointment. As a result of the full effect model developed using multiway frequency analysis, there is no significant relationship among the appointment demand type, patient behavior, and reason for the appointment variables. Based on the second order effect, each clinic experienced that operative requests are significantly higher for individual rather than batch appointments. Based on the logistic regression model, equations were generated to determine which variables contribute to the probability of patients meeting their scheduled appointment. In fact, it is not necessarily if the patient is scheduled within in the batch, but how many patients that are scheduled within the family or group.

Next, a decision-making model is developed to assist scheduling coordinators in determining when to accept or reject batch appointment request under an open-access scheduling system. Therefore, a discrete-time, discrete-space, stationary infinite-horizon Markov decision process (MDP) model was constructed to study the impact that batch arrivals have on this scheduling system. This research was motivated by a pediatric dental clinic who was interested in transitioning from a traditional scheduling system to an open-access scheduling system. However, due to unforeseen events, the actual implementation of open-access scheduling was never completed. Therefore, the MDP model presented here serves as a theoretical guide to how batch appointments influences the clinic's profitability and physician's utilization at various prescheduled appointment ratios. The model also explores the use of overbooking in order to increase the likelihood that batch appointment requests are accepted.

Although batched patients were accepted when overbooking is not allowed, the MDP model is only able to serve up to the number of physicians. In a practical setting, this implies that the physicians will have to work overtime in the event that patients are still waiting in the backlog at the end of the period. This demonstrates the importance of considering the use of overbooking in the initial phases of planning rather than requiring physicians to work overtime unexpectedly. In general, the optimal scheduling rule obtained from the MDP model suggests the batch appointment requests should always be accepted regardless of the number of patients prescheduled and the maximum overbooking limit, when the probability of no-show/cancellation for each prescheduled appointment is greater than or equal to 0.50. The clinic should always reject a request for batch appointment when the probability of no-show is less than or equal to 0.10 and the number of patients either with a physician and/or waiting in the backlog is greater than the number of physicians in the clinic. However, in spite of the probability of no-show, batch appointment requests are accepted when the number of patients either with a physician and/or waiting in the backlog is less than the number of physicians in the clinic.

Based on the information presented in the case study and MDP model, this research was extended to examine the influences of scheduling both independent and dependent demand arrivals. In the final model, a finite-horizon stochastic dynamic programming model was developed to study the impact of these appointment demand types on the clinic's profitability and the physicians' productivity. Again, this work demonstrated the use of overbooking to mitigate the risk associated with patients being dependent among each other in terms of their arrival and probability of no-show. However, unlike the MDP model, the acceptance of appointment request is restricted to the number of patients that are carried over from period to period and not prescheduled patients. This assumption is more aligned with the actual behavior of clinical environments, where patients may not be overbooked in every period.

The results of the optimal policies indicate that when the scheduling coordinator is certain that prescheduled patients will fail to meet their appointment, then the acceptance of overbooked patients is strictly based on the number of patients that are currently waiting in the backlog. However, if prescheduled patients are known to show up as planned, then the acceptance of overbooked patients is based on the total number of patients in the system (i.e. prescheduled and waiting). In fact, the optimal policy determined from the model suggests that rejections only occur when the total number of patients in the system is greater than or equal to the number of physicians in the clinic. This is similar to the results found in the MDP model.

Due to the uncertainty of patients meeting their scheduled appointment, batched patients are overbooked when the total number of patients in the system is less than the number of physicians or when the number of waiting patients is less than the batch size. This is due to the influx in demand which requires additional resources to be needed in order to ensure that the entire group will be served. However, by accepting batch appointments as overbooked rather than prescheduled patients, helps mitigate the possibility of these patients failing to meet their scheduled appointments. This, in turn, increases the productivity of the physicians but it decreases the total expected profit since the physicians may have to work overtime to ensure every patient is served.

In addition, clinics can remain profitable given the variation of probability of no-show for each of the appointment demand types when overbooking is allowed. However, in terms of the clinic's performance measures, the impact of the variability of individual patients yields better results than that observed by batched patients. Therefore, when considering multiple appointment demand types, the scheduling coordinator must optimize their ability to schedule each of the demand types in a manner that increases both the clinic's profitability and physicians' productivity.

Finally, this work has demonstrated both the need to consider independent and dependent patients and the benefits of overbooking. However, clinics must determine the optimal number of patients that they are willing to allow to wait in the backlog to be seen by an idle physician and the actual value of patient dissatisfaction due to waiting. The results of the statistical models can assist the scheduling coordinators at each clinic in determining which batch appointment requests to accept given the reason for the appointment and probability of no-show. Based on the clinical data, we determined that the probability of no-show is higher for those patients requesting batch appointments versus individual requests. As a result, clinics must understand the history of each family not meeting their scheduled appointment prior to accepting their request for multiple appointments. This will help the clinics manage the risk of scheduling batch appointments.

Future research areas include determining the actual optimal scheduling rule based on various batch sizes and examining the allocation of patients to physicians within the scheduling paradigm. However, it is also important to explore the impact of the patient's demographics on no-show rates and batch appointment requests. The demographics will consist of the patient's ethnicity, income level, insurance provider, family size, single parent or both parents, etc. By adding these demographic indicators, healthcare providers are able to further investigate the difference between clinic types and family sizes. Additional research in this area can also consider the best approach to balance the needs of batched patients and the healthcare providers. The model can also be extended to explore the effects of when only one appointment demand type can be overbooked in each period. With the change in scheduling approaches, we hope to gain further insight of the impact of dependent demand arrivals.

REFERENCES

[1] Centers for Medicare and Medicaid Services, *National Health Expenditure Data*. January 20, 2011. [Online] Available: http://www.cms.hhs.gov/NationalHealthExpendData/02_NationalHealthAccountsHistorical.asp#TopOfPage. [Accessed: March 18, 2011]

[2] I. Z. Karaesmen and I. Nakshin, "Applying pricing and revenue management in US hospitals—New perspectives," *Journal of Revenue and Pricing Management,* vol. 6, pp. 256-259, 2007.

[3] E. B. Ferlie and S. M. Shortell, "Improving the Quality of Health Care in the United Kingdom and the United States: A Framework for Change," *Milbank Quarterly,* vol. 79, pp. 281-315, 2001.

[4] J. Currie*, et al.*, *Medical Care for Children: Public Insurance, Private Insurance and Racial Differences in Utilization*: RAND, 1995.

[5] P. Reid*, et al.*, *Building a better delivery system: a new engineering/health care partnership*: National Academies Press, 2005.

[6] X. Qu*, et al.*, "Matching daily healthcare provider capacity to demand in advanced access scheduling systems," *European Journal of Operational Research,* vol. 183, pp. 812-826, 2007.

[7] M. Ross and K. Patrick, "Putting Patients First: Developing and Maintaining Patient-Centered Medical Practices," *Brookings,* 2007.

[8] D. Gupta*, et al.*, "Performance metrics for advanced access," *Journal of Healthcare Management,* vol. 51, pp. 246-258, 2006.

[9] C. D. O'Hare and J. Corlett, "The outcomes of open-access scheduling," *Family Practice Management,* vol. 11, pp. 35-38, 2004.

[10] M. B. Pinto, *et al.*, "Selling Open Access Health Care Delivery to Patients and Administrators: What's the Hook?," *Health Marketing Quartley,* vol. 19, pp. 57-70, 2002.

[11] K. Pak and N. Piersma, "Overview of OR techniques for airline revenue management," *Statistica Neerlandica,* vol. 56, pp. 480-496, 2002.

[12] L. R. Weatherford, *et al.*, "Modeling the customer arrival process and comparing decision rules in perishable asset revenue management situations," *Transportation Science,* vol. 27, p. 239, 1993.

[13] J. A. Fitzsimmons and M. J. Fitzsimmons, *Service Management: Operations, Strategy, and Information Technology*, 5th ed. New York: McGraw-Hill/Irwin Education, 2006.

[14] K. Talluri and G. van Ryzin, *The theory and practice of revenue management*. New York: Kluwer Academic Publishers, 2004.

[15] D. Gupta and L. Wang, "Revenue management for a primary-care clinic in the presence of patient choice," *Operations Research,* vol. 56, pp. 576-592, 2008.

[16] M. L. Brandeau, *et al.*, *Operations research and health care. A Handbook of methods and applications* vol. 70: International Series in Operations Research & Management Science, 2004.

[17] V. L. Smith-Daniels, *et al.*, "Capacity Management in Health Care Services: Review and Future Research Directions," *Decision Sciences,* vol. 19, pp. 889-919, 1988.

[18] J. I. McGill and G. J. Van Ryzin, "Revenue Management: Research Overview and Prospects," *Transportation Science,* vol. 33, pp. 233-256, 1999.

[19] W. C. Chiang, *et al.*, "An overview of research on revenue management: current issues and future research," *International Journal of Revenue Management,* vol. 1, pp. 97-128, 2007.

[20] D. Pencheon, "Managing demand: Matching demand and supply fairly and efficiently," *BMJ: British Medical Journal,* vol. 316, p. 1665, 1998.

[21] S. Chopra and P. Meindl, *Supply Chain Management: Strategy, Planning, and Operation*, 2nd ed. Upper Saddle River: Pearson Prentice Hall, 2003.

[22] M. Murray and C. Tantau, "Same-day appointments: exploding the access paradigm," *Family Practice Management,* vol. 7, pp. 45-50, 2000.

[23] L. Schneck, "Open-access scheduling can open the door to better performance," *Medical Group Management Association Connexion,* vol. 1, pp. 39-40, 2001.

[24] M. Murray, "Answers to your questions about same-day scheduling," *Family Practice Management,* vol. 12, pp. 59-64, 2005.

[25] S. Herriott, "Reducing Delays and Waiting Times with Open-Access Scheduling," *Family Practice Management,* pp. 38–46, 1999.

[26] D. Gupta and B. Denton, "Appointment scheduling in health care: Challenges and opportunities," *IIE Transactions,* vol. 40, pp. 800-819, 2008.

[27] S. Savin, "Managing Patient Appointments in Primary Care," *International Series in Operations Research & Management Science* vol. 91, pp. 123-150, 2006.

[28] L. V. Green*, et al.*, "Providing timely access to care: what is the right patient panel size?," *Joint Commission Journal on Quality and Patient Safety,* vol. 33, pp. 211-218, 2007.

[29] M. Murray*, et al.*, "Panel Size: How Many Patients Can One Doctor Manage?," *Family Practice Management,* vol. 14, p. 44, 2007.

[30] C. M. Kilo and S. Endsley, "As good as it could get: remaking the medical practice," *Family Practice Management,* vol. 7, pp. 48-52, 2000.

[31] M. Murray and D. M. Berwick, "Advanced access - Reducing waiting and delays in primary care," *Journal of the American Medical Association,* vol. 289, pp. 1035-1040, 2003.

[32] G. D. Randolph, "Where next for advanced access: will it be embraced by specialties?," *NC Medical Journal,* vol. 66, pp. 226-228, 2005.

[33] M. Miller, "Open Access Scheduling: Examining Its Relationship to Patient Satisfaction," *Group Practice Journal* vol. 56, p. 37, 2007.

[34] M. Murray*, et al.*, "Improving Timely Access to Primary Care Case Studies of the Advanced Access Model," *Journal of the American Medical Association,* vol. 289, pp. 1042-1046, 2003.

[35] D. G. Bundy*, et al.*, "Open access in primary care: Results of a North Carolina Pilot Project," *Pediatrics,* vol. 116, pp. 82-87, 2005.

[36] A. Knight*, et al.*, "Reduced waiting times for the GP: two examples of" advanced access" in Australia," *Medical Journal of Australia,* vol. 183, p. 101, 2005.

[37] M. Meyers, "Changing business practices for appointing in military outpatient medical clinics: the case for a true" open access" appointment scheme for primary care," *Journal of Healthcare Management,* vol. 48, pp. 125-139, 2003.

[38] B. Armstrong*, et al.*, "Reinventing veterans health administration: Focus on primary care," *Journal of Healthcare Management,* vol. 50, pp. 399-408, 2005.

[39] J. G. Kennedy and J. T. Hsu, "Implementation of an Open Access Scheduling System in a Residency Training Program," *Family Medicine* vol. 35, pp. 666-670, 2003.

[40] J. R. Steinbauer*, et al.*, "Implementing Open-Access Scheduling in an Academic Practice," *Family Practice Management,* vol. 13, p. 59, 2006.

[41] J. S. Gill, "A nonfinancial approach to financial improvement of medical groups through advanced access," *Journal of Healthcare Management,* vol. 49, pp. 271-7, 2004.

[42] F. Belardi*, et al.*, "A controlled trial of an advanced access appointment system in a residency family medicine center," *Family Medicine,* vol. 36, pp. 341-345, 2004.

[43] E. Newman*, et al.*, ""The rheumatologist can see you now": successful implementation of an advanced access model in a rheumatology practice," *Arthritis & Rheumatism,* vol. 51, 2004.

[44] L. Solberg*, et al.*, "Key issues in transforming health care organizations for quality: the case of advanced access," *Joint Commission Journal on Quality and Patient Safety,* vol. 30, p. 15, 2004.

[45] K. Terry, "Don't be afraid of same-day scheduling," *Medical Economics,* vol. 81, pp. 64-6, 2004.

[46] A. Mehrotra*, et al.*, "Implementing Open-Access Scheduling of Visits in Primary Care Practices: A Cautionary Tale," *Annals of internal medicine,* vol. 148, pp. 915-W201, 2008.

[47] S. Ahluwalia and M. Offredy, "A qualitative study of the impact of the implementation of advanced access in primary healthcare on the working lives of general practice staff," *BMC Family Practice,* vol. 6, p. 39, 2005.

[48] M. Pickin*, et al.*, "Evaluation of advanced access in the national primary care collaborative," *The British Journal of General Practice,* vol. 54, p. 334, 2004.

[49] D. H. Parente*, et al.*, "A Pre-Post Comparison of Service Operational Efficiency and Patient Satisfaction Under Open Access Scheduling," *Health Care Management Review,* vol. 30, p. 220, 2005.

[50] K. J. Bennett and E. G. Baxley, "The Effect of a Carve-out Advanced Access Scheduling System on No-show Rates," *Family Medicine,* vol. 41, pp. 51-56, 2009.

[51] S. Lewandowski, *et al.*, "Increasing Primary Care Physician Productivity: A Case Study," *American Journal of Managed Care,* vol. 12, pp. 573-576, 2006.

[52] M. E. O'Connor, *et al.*, "Effect of open access scheduling on missed appointments, immunizations, and continuity of care for infant well-child care visits," *Archives of Pediatrics & Adolescent Medicine,* vol. 160, pp. 889-893, 2006.

[53] M. Fine and S. Busselen, "Would You Like to See a Doctor Today?'," *Family Practice Management,* vol. 15, p. 7, 2008.

[54] J. Sperl-Hillen, *et al.*, "Peer Reviewed: The Effect of Advanced Access Implementation on Quality of Diabetes Care," *Preventing Chronic Disease,* vol. 5, 2008.

[55] G. D. Randolph, *et al.*, "Effect of NC Open Access Collaborative on Medicaid Utilization and Costs for Children in Primary Care Practices," ed.

[56] P. DeLaurentis, *et al.*, "Open access appointment scheduling–an experience at a community clinic," in *IIE Annual Conference and Expo*, 2006.

[57] R. E. Giachetti, *et al.*, "Assessing the viability of an open access policy in an outpatient clinic: a discrete-event and continuous simulation modeling approach," 2005, pp. 2246-2255.

[58] R. Kopach, *et al.*, "Effects of clinical characteristics on successful open access scheduling," *Health Care Management Science,* vol. 10, pp. 111-124, 2007.

[59] K. Muthuraman and M. Lawley, "A stochastic overbooking model for outpatient clinical scheduling with no-shows," *IIE Transactions,* vol. 40, pp. 820-837, 2008.

[60] X. Qu and J. Shi, "Effect of two-level provider capacities on the performance of open access clinics," *Health Care Management Science,* vol. 12, pp. 99-114, 2009.

[61] K. Talluri and G. van Ryzin, "Revenue Management Under a General Discrete Choice Model of Consumer Behavior," *Management Science,* vol. 50, pp. 15-33, 2004.

[62] M. Emeksiz*, et al.*, "A yield management model for five-star hotels: Computerized and non-computerized implementation," *International Journal of Hospitality Management,* vol. 25, pp. 536-551, 2006.

[63] T. Koide and H. Ishii, "The hotel yield management with two types of room prices, overbooking and cancellations," *International Journal of Production Economics,* vol. 93, pp. 417-428, 2005.

[64] C. K. Anderson*, et al.*, "Revenue management: A real options approach," *Naval Research Logistics,* vol. 51, pp. 686-703, 2004.

[65] K. Amaruchkul*, et al.*, "Single-Leg Air-Cargo Revenue Management," *Transportation Science* vol. 41, pp. 457-469, 2007.

[66] K. Lai*, et al.*, "A stochastic approach to professional services firms' revenue optimization," *European Journal of Operational Research,* vol. 182, pp. 971-982, 2007.

[67] R. Metters and V. Vargas, "Yield management for the nonprofit sector," *Journal of Service Research,* vol. 1, pp. 215-226, 1999.

[68] J. Pinder, "Using revenue management to improve pricing and capacity management in programme management," *Journal of the Operational Research Society,* vol. 56, pp. 75-87, 2004.

[69] S. E. Kimes and G. M. Thompson, "An evaluation of heuristic methods for determining the best table mix in full-service restaurants," *Journal of Operations Management,* vol. 23, pp. 599-617, 2005.

[70] D. Bertsimas and R. Shioda, "Restaurant Revenue Management," *Operations Research,* vol. 51, pp. 472-486, 2003.

[71] F. H. d. Harris and J. Pinder, "A revenue management approach to demand management and order booking in assemble-to-order manufacturing," *Journal of Operations Management,* vol. 13, pp. 299-309, 1995.

[72] C. Maglaras, "Revenue management for a multi-class single-server queue," *Operations Research,* vol. 54, pp. 914-932, 2004.

[73] R. Quante*, et al.*, "A Stochastic Dynamic Programming Approach to Revenue Management in a Make-to-Stock Production System," *Working paper*.

[74] R. Quante*, et al.*, "Revenue management and demand fulfillment: matching applications, models, and software," *OR Spectrum,* vol. 31, pp. 31-62, 2009.

[75] S. Kim and R. E. Giachetti, "A Stochastic Mathematical Appointment Overbooking Model for Healthcare Providers to Improve Profits," *IEEE Transactions on Systems, Man & Cybernetics: Part A,* vol. 36, pp. 1211-1219, 2006.

[76] L. Qi and Z. Yan, "Capacity Control of Revenue Management in Wards of Community Hospital," in *Service Systems and Service Management, 2006 International Conference* 2006, pp. 368-372.

[77] D. Gupta, "Capacity planning and patient flow management," *Operations Research for Health Care Delivery Systems, World Technology Evaluation Center, Inc., Baltimore, Maryland,* 2005.

[78] N. T. J. Bailey, "A study of queues and appointment systems in hospital out-patient departments, with special reference to waiting-times," *Journal of the Royal Statistical Society. Series B (Methodological),* pp. 185-199, 1952.

[79] W. P. Pierskalla and D. J. Brailer, "Applications of operations research in health care delivery," in *Operations Research and the Public Sector*. vol. 6, ed: Elsevier Science, 1994, pp. 469–505.

[80] P. R. Harper and H. M. Gamlin, "Reduced outpatient waiting times with improved appointment scheduling: a simulation modelling approach," *OR Spectrum,* vol. 25, pp. 207-222, 2003.

[81] K. J. Klassen and T. R. Rohleder, "Scheduling outpatient appointments in a dynamic environment," *Journal of Operations Management,* vol. 14, pp. 83-101, 1996.

[82] T. Cayirli and E. Veral, "Outpatient Scheduling in Health Care: A Review of Literature," *Production and Operations Management,* vol. 12, pp. 519-549, 2003.

[83] W. England and S. D. Roberts, "Applications of computer simulation in health care," 1978, pp. 665-677.

[84] G. Kaandorp and G. Koole, "Optimal outpatient appointment scheduling," *Health Care Management Science,* vol. 10, pp. 217-229, 2007.

[85] R. Giachetti, "A simulation study of interventions to reduce appointment lead-time and patient no-show rate," 2008, pp. 1463-1468.

[86] L. R. LaGanga and S. R. Lawrence, "Clinic Overbooking to Improve Patient Access and Increase Provider Productivity," *Decision Sciences,* vol. 38, pp. 251-276, 2007.

[87] B. J. Johnson*, et al.*, "Reduction and management of no-shows by family medicine residency practice exemplars," *Annals of Family Medicine* vol. 5, pp. 534-539, 2007.

[88] S. N. Forjuoh*, et al.*, "Open-access appointment scheduling in family practice: comparison of a demand prediction grid with actual appointments," *The Journal of the American Board of Family Medicine,* vol. 14, p. 259, 2001.

[89] B. Zeng*, et al.*, "Clinic Scheduling Models with Overbooking for Patients with Heterogeneous No-Show Probabilities," *Working paper.,* 2008.

[90] L. R. LaGanga and S. R. Lawrence, "Appointment Scheduling with Overbooking to Mitigate Prodcutivity Loss From No-Shows," *Working paper.,* 2009.

[91] N. Liu, *et al.*, "Dynamic scheduling of outpatient appointments under patient no-shows and cancellations," *Manufacturing & Service Operations Management,* vol. 12, pp. 347-364, 2010.

[92] B. Zeng, *et al.*, "Clinic Overbooking and Patient Response: A Game Theoretic Approach," *IIE Annual Conference and Expo,* 2009.

[93] R. E. Giachetti, "A simulation study of interventions to reduce appointment lead-time and patient no-show rate," 2008, pp. 1463-1468.

[94] United States Census Bureau, *America's Families and Living Arrangements:* January 14, 2010. [Online] Available: http://www.census.gov/population/www/socdemo/hh-fam/cps2009.html. [Accessed: January 23, 2010]

[95] H. Abdus-Salaam, *et al.*, "Modeling dependent demand arrivals within an open-access scheduling system," in *Science and Technology for Humanity (TIC-STH), 2009 IEEE Toronto International Conference* 2010, pp. 256-261.

[96] N. L. Lacy, *et al.*, "Why We Don't Come: Patient Perceptions on No-Shows," *The Annals of Family Medicine,* vol. 2, pp. 541-545, 2004.

[97] X. Qu and R. Rardin, "A statistical model for the prediction of patient non-attendance in a primary care clinic," *RCHE Publications,* p. 10, 2006.

[98] C. G. Moore, *et al.*, "Time and Money: Effects of No-Shows at a Family Practice Residency Clinic," *Family Medicine,* vol. 33, pp. 522-527, 2001.

[99] H. Dove and K. Schneider, "The Usefulness of Patients' Individual Characteristics in Predicting No-Shows in Outpatient Clinics," *Medical Care,* vol. 19, pp. 734-740, 1981.

[100] B. G. Tabachnick and L. S. Fidell, *Using Multivariate Statistics*, 5th ed. New York: Pearson Education, Inc. Allyn and Bacon, 2007.

[101] H. Abdus-Salaam and L. Davis, "A Comparison of Patient Dependency at Public and Private Pediatric Clinics: A Case Study " *Working paper.,* 2010.

[102] D. Bundy and G. Randolph, "Open Access Implementation and Diabetes Outcomes," *Journal of General Internal Medicine,* vol. 24, pp. 991-991, 2009.

[103] P. DeLaurentis*, et al.*, "A configurable framework for open access scheduling with continuous improvement in outpatient clinics," Working paper, 2006.

[104] J. Kros*, et al.*, "Overbooking Increases Patient Access at East Carolina University's Student Health Services Clinic," *Interfaces,* vol. 39, pp. 271-287, 2009.

[105] L. Robinson and R. Chen, "A Comparison of Traditional and Open-Access Policies for Appointment Scheduling," *Manufacturing & Service Operations Management,* vol. 12, pp. 330-346, 2009.

[106] L. LaGanga and S. Lawrence, "Yield Management In Health Care Clinics With Walk-In Traffic," *Leeds School of Business Research Paper Series,* 2008.

[107] R. Howard, *Dynamic Programming and Markov Processes.* Cambridge, MA: Technology Press–Wiley, 1960.

[108] S. Chakraborty*, et al.*, "Sequential Clinical Scheduling with Patient No-shows and General Service Time Distributions," *IIE Transactions,* vol. 42, pp. 354-366, 2010.

[109] W. Winston, *Operations Research: Applications and Algorithms*, 2nd ed. Boston, MA: PWS-Kent Pub. Co., 1991.

[110] D. Bertsekas, *Dynamic Programming and Optimal Control*. Belmont, MA: Athena Scientific, 2005.

APPENDIX A

SAS Code

Multiway Frequency Analysis

```
data reading;
input batch$ status$ reason$ freq;
cards;

[ACTUAL DATA]
;

ods rtf;
proc catmod;
weight freq;
model batch*status*reason=_response_/noiter ;
loglin batch|status|reason;
run;

proc freq;
tables batch*status*reason batch*status batch*reason status*reason
batch status reason/ all;
weight freq;
run;

ods rtf close;
```

Logistic Regression

```
data broken;
input day status provider reason duration size batch;

if status=1 then sched=1; else sched=0;
if status=2 then noshow=1; else noshow=0;

if day=1 then monday=1; else monday=0;
if day=2 then tuesday=1; else tuesday=0;
if day=3 then wednesday=1; else wednesday=0;
if day=4 then thursday=1; else thursday=0;

if provider=1 then dentist=1; else dentist=0;
if provider=2 then hygienist=1; else hygienist=0;

if reason=1 then recall=1; else recall=0;
if reason=2 then operative=1; else operative=0;

if duration=1 then lessorequal20=1; else lessorequal20=0;
if duration=2 then equal30=1; else equal30=0;
if duration=3 then equal40=1; else equal40=0;
if duration=4 then equal50=1; else equal50=0;
if duration=5 then equal60=1; else equal60=0;
```

```
if size=1 then size1=1; else size1=0;
if size=2 then size2=1; else size2=0;
if size=3 then size3=1; else size3=0;
if size=4 then size4=1; else size4=0;
if size=5 then size5=1; else size5=0;

cards;

[ACTUAL DATA]
;

ods rtf;
proc logistic descending;
model   sched noshow= monday tuesday wednesday thursday dentist
hygienist recall operative lessorequal20 equal30 equal40 equal50
equal60 size1 size2 size3 size4 size5 batch/selection=stepwise;
run;

ods rtf close;
```

APPENDIX B

Matlab Code for MDP

Policy iteration code

```
function [export_matrix]=Policy_iter6()
%%%

clear;
clc;

format short;

%Input parameters
MaxB = 0;
Cp = 3;
BS = 2;

revenue_patient = 135;
penalty_backlog = 54;
penalty_unutilized = 13.5;

total_column = 14+2*(MaxB+1)*(Cp+1);
export_matrix = zeros(1,total_column);

    for prob_ns_iteraction=1:11  %%%
    switch prob_ns_iteraction
        case 1
            prob_ns = 0;
        case 2
            prob_ns = 0.1;
        case 3
            prob_ns = 0.2;
        case 4
            prob_ns = 0.3;
        case 5
            prob_ns = 0.4;
        case 6
            prob_ns = 0.5;
        case 7
            prob_ns = 0.6;
        case 8
            prob_ns = 0.7;
        case 9
            prob_ns = 0.8;
        case 10
            prob_ns = 0.9;
        case 11
            prob_ns = 1;
    end;
```

```
for prob_BA_iteration=1:11
    switch prob_BA_iteration
        case 1
            prob_BA = 0;
        case 2
            prob_BA = 0.1;
        case 3
            prob_BA = 0.2;
        case 4
            prob_BA = 0.3;
        case 5
            prob_BA = 0.4;
        case 6
            prob_BA = 0.5;
        case 7
            prob_BA = 0.6;
        case 8
            prob_BA = 0.7;
        case 9
            prob_BA = 0.8;
        case 10
            prob_BA = 0.9;
        case 11
            prob_BA = 1;
    end;

[p_full,busy_dist,q,noshow_prob,state1]=ClinicSched3(MaxB,Cp,revenue_pa
tient,penalty_backlog,penalty_unutilized,prob_ns,prob_BA, BS);
     %function [pi,P]=ClinicSched(MaxB,Cp)
     %Cp = Maximum number of physicians
     %MaxB = Maximum # of patients that can be in backlog
     %prob_ns = noshow_prob;

     [size_p_full_x,size_p_full_y,size_p_full_d] = size(p_full);
     size_p = size(p_full,2);

     %%%%%%% Policy Iteration
     %%%%% Generate the first policy [1 1 1 ....]
     for i=1:size_p
         d(i,1)=1;
     end;

     %%%%%% Calculate the q_full for the test step
     % for k=1:size_p_full_d
     %     for i=1:size_p
     %         sum_pr = 0;
     %         for j=1:size_p
     %             sum_pr = sum_pr + (p_full(i,j,k)*r_full(i,j,k));
     %         end;
     %         q(i,k)=sum_pr;
     %     end;
     % end;
```

```
gain = -9999;
for a=1:100
    %%%%%%% Beginning of the algorithm
    results = zeros(1,5);

    p = zeros(size_p,size_p);
    for k=1:size_p
        p(k,:)=p_full(k,:,d(k));
        %r(k,:)=r_full(k,:,d(k));
    end;

    %Calculate steady-state probabilities
    I = eye(size_p);
    A = (I - p);
    A(:,size_p) = 1;


   %b treatment
   b = zeros(size_p,1);
   b(size_p,1) = 1;
   bT = b';

   %steady state pi vector
   pi = bT*inv(A);

    x = inv(A)*q;
    new_gain = x(size_p,1);

    v=zeros(size_p,1);
    for i=1:size_p-1
        v(i,1)=x(i,1);
    end;

    %%%% Test step
    for i=1:size_p
        for k=1:size_p_full_d
            sum_pv=0;
            for j=1:size_p
                sum_pv = sum_pv + (p_full(i,j,k)*v(j));
            end;
            results = [results;i k q(i,k) sum_pv q(i,k)+sum_pv];
        end;
    end;
    %%%%% Exclude the first zeros matrix line
    if sum(results(1,:)==0)
        results(1,:) = [];
    end;

    temp=zeros(size_p_full_d,1);
    for i=0:size_p-1
        for n=1:size_p_full_d
```

```
                temp(n,1)=results(2*i + n,5);
            end;
            [value,index] = max(temp(:,1));
            new_policy(i+1,1) = index;
        end;

        disp('iteration');
        disp(a);

        disp('policy');
        disp(d);

        disp('gain');
        disp(new_gain);

        disp('v');
        disp(v);

        if new_gain>gain
            d = new_policy;
            gain = new_gain;
        else
            %pause;
            break;
        end;
    end;
    disp('Interaction Parameters');
    disp('Revenue');
    disp(revenue_patient);
    disp('BackLog');
    disp(penalty_backlog);
    disp('Unutilized');
    disp(penalty_unutilized);
    disp('Probability of No-show');
    disp(prob_ns);

%%E[PS|i] =>Expected prescheduled served given i
%%E[BS|j] =>Expected backlogged served given j
s=zeros(1,10);
k=1;
expected_presched_served=0;
expected_backlog_served=0;
 for i=0:Cp
    for j=0:MaxB
        %%%%%%%%%%
        %define pdf for no shows
        %%%%%%%%%%%%%%%%%%%
        E_ps_i=0; %expected prescheduled served given i
        E_bs_i=0; % expected backlogged served given i
        ns_pdf=zeros(1,3);
        for t=0:i
            ns_pdf(t+1)=nchoosek(i,t)*(prob_ns^t)*(1-prob_ns)^(i-t);
```

```
        % ns_pdf(t+1)=nchoosek(i,t)*((1-prob_ns)^t)*((prob_ns)^(i-t));
        end
        y=ns_pdf;
        %%%%%%%%%%%%%%%%%%%%%%%%%
        s(k,1)=i;
        s(k,2)=j;
        state9= encode_state(i, j, MaxB+1);
        for ns=1:i+1
            x = 1:ns;
            %y = binopdf(x,i+1,prob_ns);
            E_ps_i = E_ps_i + ((i- (ns-1))*y(ns));
            E_bs_i= E_bs_i +  min (j,Cp-(i-(ns-1)) )*y(ns)*prob_BA +
min (j+BS, Cp-(i-(ns-1))  )*y(ns)*(1-prob_BA);
        end;
        s(k,3) = E_bs_i; %expected_backlog_served;
        s(k,4) = E_ps_i; %expected_presched_served;
        s(k,5) = s(k,3)+s(k,4); %(expected_backlog_served +
expected_presched_served); %Expected total served
%          s(k,6) = pi(3*i+j+1);
        s(k,6) = pi(state9+1);  %steady state probability
        s(k,7) = s(k,5)*s(k,6); %expected total served
        s(k,8) = s(k,2)*s(k,6); %expected patients assigned to backlog
        s(k,9) = s(k,3)*s(k,6);  %expected backlog served
        s(k,10) = s(k,4)*s(k,6); %expected prescheduled served
        expected_presched_served = expected_presched_served + s(k,10);
        expected_backlog_served = expected_backlog_served + s(k,9);
        k=k+1;
    end;
 end;

%%E[S|(i,j)] =>E[PS|i]+E[BS|j]
% expected_served=zeros(1,((MaxB+1)*(Cp+1)));
% for i=1:k
%     expected_served=expected_served + expected_presched_served +
expected_backlog_served;
% end;
% expected_served

%%E[S] => sumproduct of E[S|(i,j)] and pi(i,j)
total_expected_served=0;
total_expected_served= sum(s(:,7));
total_expected_served;

%Utilization of physicians
Utilization=0;
Utilization= (total_expected_served/Cp)*100;
Utilization;

%Expected number in backlog
expected_backlog=0;
expected_backlog= sum(s(:,8));
expected_backlog;
```

```
%Expected number in backlogged patients served
exp_backlog_served=0;
exp_backlog_served= sum(s(:,9));
exp_backlog_served;

%Expected number in prescheduled patients served
exp_presch_served=0;
exp_presch_served= sum(s(:,10));
exp_presch_served;

    export_matrix = [export_matrix;MaxB Cp BS revenue_patient
penalty_backlog penalty_unutilized prob_ns prob_BA new_gain
total_expected_served Utilization expected_backlog exp_backlog_served
exp_presch_served d' pi];
    save my_data.out export_matrix -ASCII
%
% end;
% end;
% end;
end;
end;
%/////////////////////////////////////////////////////////
%/*  function: decode_state */
%//  convert state s to values Is and Ir
%/////////////////////////////////////////////////////////
function [Is,Ir]=decode_state(s, M2)

    if (s == 0)

        Is = 0;
        Ir = 0;

    else

        Is =floor(s/M2);
        Ir = mod(s,M2);
    end


%/////////////////////////////////////////////////////////
%/*  function: encode_state */
%//  convert state s to values I1 and I2
%/////////////////////////////////////////////////////////
function [state] = encode_state(I1, I2,  M2)
 state = I1*(M2) + I2;
```

Q-Matrix

```
function
[P,busy_dist,q,noshow_prob,state1]=ClinicSched3(MaxB,Cp,parameter1,para
meter2,parameter3,parameter4,parameter5,parameter6)
%Cp = Maximum number of physicians
%MaxB = Maximum # of patients that can be in backlog

%Model Assumptions
%1. Assume request for batch appointment that is made in the beginning
of the
%period can be processed in the same period if there is a no-show.
%2. Assume pre-schedule appointments have priority over same day
appointments
%3. Assume no show distribution is conditioned on the number of
prescheduled
%appointments and follows a binomial distribution with success
probability
%p_ns
%4. Open slots are perishable and can't be carried forward into the
next
%period.
%5. Probability that a prescheduled appointment exists in the next
period
%is defined by a binomial distribution with success_prob=busy_prob.

p_ns=parameter4; %probability a person arrives therefore probabilty
they dont is 1-p_ns
busy_prob= 0.5; %probability of preschedule appointment
busy_dist=zeros(1,Cp+1);
Numstates=(MaxB+1)*(Cp+1);  %size of Pmatrix
%B_arrivals = [0.3 0.7];  %probability of batch arrival of 0 or 2
B_arrivals = [parameter5 1-parameter5];  %probability of batch arrival
of 0 or b
P=zeros(Numstates,Numstates,2);
fixed_cost=0;
BS=parameter6; %%%Batch Size Cp/2 rounded up

%alternative 1 is accept alternative 2 is reject
q=zeros(Numstates,2);
%build busy_prob distribution
for k=0:Cp
    busy_dist(k+1)=nchoosek(Cp,k)*(busy_prob^k)*(1-busy_prob)^(Cp-k);
%binomial distribution
end


%build P-Matrix for accept alternative
for s1=1:Cp
    for s2=0:MaxB     %can't accept backlog if MaxB is reached
        state1=encode_state(s1,s2,MaxB+1);
        for NS=0:s1   %iterate over number of noshows
```

```
            noshow_prob = nchoosek(s1,NS)*(p_ns^NS)*(1-p_ns)^(s1-NS);
%binomial distribution
%              noshow_prob = nchoosek(s1,NS)*(1-p_ns^NS)*(p_ns)^(s1-NS);
%binomial distribution
            for a=0:1  %iterate over number of batch arrivals
                for b=0:Cp   %iterate over # of presched
                    new_s1=b;
                    Numavail = Cp-(s1-NS); %num available slots
                    Numreqd= s2+a*BS;      %num people in backlog
                    if (Numreqd < Numavail)
                        new_s2=0;
                    else
                        new_s2=min(Numreqd-Numavail,MaxB); %can't
exceed backlog.
                    end
                        %make sure a very high penalty is assessed so
that
                        %this case is rejected
                     state2=encode_state(new_s1,new_s2,MaxB+1);
                    P(state1+1,state2+1,1)=P(state1+1,state2+1,1)+
busy_dist(b+1)*noshow_prob*B_arrivals(a+1);
                    %calculate reward
                    BatchServed=min(Numreqd,Numavail);
                    SchedServed= s1-NS;
                    BatchRemain= s2+ BS*a-BatchServed;
                    Numidle= max(Numavail-Numreqd,0);
                    Revenue = parameter1*(BatchServed+SchedServed)-
parameter2*BatchRemain - parameter3*Numidle - fixed_cost;
                    q(state1+1,1)=q(state1+1,1)+
Revenue*busy_dist(b+1)*noshow_prob*B_arrivals(a+1);
                  %calculate new s2 for reject decision
                  if (s2 < Numavail) new_s2r=0;
                  else
                      new_s2r=max(s2-Numavail,0);
                  end
                  state2_reject=encode_state(new_s1,new_s2r,MaxB+1);

P(state1+1,state2_reject+1,2)=P(state1+1,state2_reject+1,2)+
busy_dist(b+1)*B_arrivals(a+1)*noshow_prob;
                  %calculate reward for reject decision
                    BatchServed=min(s2,Numavail);
                    SchedServed= s1-NS;
                    BatchRemain= s2-BatchServed;
                    Numidle= max(Numavail-s2,0);
                    Revenue = parameter1*(BatchServed+SchedServed)-
parameter2*BatchRemain - parameter3*Numidle - fixed_cost;
                    q(state1+1,2)=q(state1+1,2)+
Revenue*busy_dist(b+1)*noshow_prob*B_arrivals(a+1);
                end
            end
        end
    end
end
```

```
%build P-matrix for accept alternative when s1=0,which implies no-
noshow
%calculations
 s1=0;
 for s2=0:MaxB     %can't accept backlog if MaxB is reached
    state1=encode_state(s1,s2,MaxB+1);
    for a=0:1  %iterate over number of batch arrivals
        for b=0:Cp   %iterate over # of presched
          new_s1=b;
          Numavail = Cp-s1; %num available slots
          Numreqd= s2+BS*a;      %num people in backlog
          if (Numreqd < Numavail)
               new_s2=0;
          else
                new_s2=min(Numreqd-Numavail,MaxB); %can't exceed
backlog.
          end
              state2=encode_state(new_s1,new_s2,MaxB+1);
              P(state1+1,state2+1,1)=P(state1+1,state2+1,1)+
busy_dist(b+1)*B_arrivals(a+1);
              %calculate reward
               BatchServed=min(Numreqd,Numavail);
               SchedServed= s1-NS;
               BatchRemain= s2+BS*a-BatchServed;
               Numidle= max(Numavail-Numreqd,0);
               Revenue = parameter1*(BatchServed+SchedServed)-
parameter2*BatchRemain - parameter3*Numidle - fixed_cost;
               q(state1+1,1)=q(state1+1,1)+
Revenue*busy_dist(b+1)*B_arrivals(a+1);
              %calculate new s2 for reject decision
              new_s2r=max(s2-Numavail,0);
              state2_reject=encode_state(new_s1,new_s2r,MaxB+1);

P(state1+1,state2_reject+1,2)=P(state1+1,state2_reject+1,2)+
busy_dist(b+1)*B_arrivals(a+1);
                  %calculate reward for reject decision
                    BatchServed=min(s2,Numavail);
                    SchedServed= s1-NS;
                    BatchRemain= s2-BatchServed;
                    Numidle= max(Numavail-s2,0);
                    Revenue = parameter1*(BatchServed+SchedServed)-
parameter2*BatchRemain - parameter3*Numidle - fixed_cost;
                    q(state1+1,2)=q(state1+1,2)+
Revenue*busy_dist(b+1)*B_arrivals(a+1);

     end
    end
end


%/////////////////////////////////////////////////////////
%/*  function: decode_state */
%//  convert state s to values Is and Ir
```

```
%//////////////////////////////////////////////////////////
function [Is,Ir]=decode_state(s, M2)

    if (s == 0)

        Is = 0;
        Ir = 0;

    else

        Is =floor(s/M2);
        Ir = mod(s,M2);
    end


%//////////////////////////////////////////////////////////
%/*  function: encode_state */
%//  convert state s to values I1 and I2
%//////////////////////////////////////////////////////////
function [state] = encode_state(I1, I2,  M2)
 state = I1*(M2) + I2;
```

<u>Decode State (Mapping Function)</u>

```
%//////////////////////////////////////////////////////////
%/*  function: decode_state */
%//  convert state s to values Is and Ir
%//////////////////////////////////////////////////////////
function [Is,Ir]=decode_state(s, M2)

    if (s == 0)

        Is = 0;
        Ir = 0;

    else

        Is =floor(s/M2);
        Ir = mod(s,M2);
    end
```

<u>Encode State (Mapping Function)</u>

```
%//////////////////////////////////////////////////////////
%/*  function: encode_state */
%//  convert state s to values I1 and I2
%//////////////////////////////////////////////////////////
function [state] = encode_state(I1, I2,  M2)
 state = I1*(M2) + I2;
```

APPENDIX C

Matlab Code for SDP

Compute State

```
function State = ComputeState(C, MaxB, BS)
%%%Generates the intial state for the SDP
%%%State=(individuals, batch, backlog)

State = [];
for i=0:C %prescheduled individuals
    for j=0:BS:BS  %prescheduled batched patients
        for k=0:MaxB %Backlogged/waiting individual patients
            for l=0:MaxB %Backlogged/waiting batch patients
                if(i+j<=C && i+j+k+l<=2*C && k+l<=MaxB)
                    State = [State;[i j k l]];
                end
            end
        end
    end
end
```

Finite-Horizon SDP

```
function export_matrix = finiteSDP_lagv2()
%FiniteSDP = finiteSDP()

clear all
clc
N = 8;    %number of periods

MaxB=3; %%%maximum number of overbooked appointments
BS=2;   %%%number of scheduled within batch
C=3;    %%%number of physicians
ukMax = 3; %%%%Also represents the highest batch size allowed
Decision = 0:ukMax;

StateTEMP = ComputeState(C, MaxB, BS); % function call for state space

[NumOfStates StateDim] = size(StateTEMP);
InitialState= StateTEMP;


load('schedule') % Load the prescheduled for experiments
StateAll = schedule;
 NumOfExperiments  = size(StateAll,1);

 prob_ns_indAll = 0:0.2:1;  %Varies the probability of no-show for each
experiment
```

```
 prob_ns_batchAll =0:0.2:1;
 batch= [0 BS];
%  StateDim = 4;

 %%%Cost Parameters
 RevenueVal = 135;
 PenUn = 13.5;
 PenOT = 54;

 for expr = 1:NumOfExperiments
     export_matrix = [];
     count = 0;  %%%keeps count of the number of experiments ran
     for prob_ns_ind_iteraction=2:2 %1:6  %%%
%           prob_ns_ind = prob_ns_indAll(prob_ns_ind_iteraction);
               prob_ns_ind = 0.2;
         for prob_ns_bat_iteraction=2:2 %1:6  %%%
%               prob_ns_batch =
prob_ns_batchAll(prob_ns_bat_iteraction);
                      prob_ns_batch = 0.2;
             ns_batch = [1-prob_ns_batch prob_ns_batch];    % batch no-
show probability

             count = count + 1;
             presched_ind = [];
             presched_bat = [];
             ns_indProb =[];
             ns_batProb =[];
             expState = [];
             tempExpectation = zeros(ukMax+1, 1); % initialiation
             Expectation  = zeros(NumOfStates, 2*N);
             opt_uk  = zeros(NumOfStates, N);


             for stage = N-1:-1:0 %0:N-1
                 State{stage+1} = InitialState;
                 if stage~= N-1 %0
                     preState = StateAll(expr,2*stage+1:2*stage+2);

                     backlog_ind = ActualOverflow(:,3);  %%%Determines
overflow patients for other stages
                     backlog_bat = ActualOverflow(:,4);
                     State{stage+1} = [repmat(preState, [NumOfStates
1]), backlog_ind , backlog_bat ];
                 end


                 ns_batProb = [];
                 ns_indProb = [];

                 % Computing the necessary probability distributions
                 for z = 1:NumOfStates
                     temp_matrix = [];
```

```
                    previousExpCost = 0;
                    ns_batProb = [];
                    ns_indProb = [];
                    presched_ind =[];
                    presched_bat = [];
                    for i= 0:C
                        ns_indProb = [];
                        ns_batProb = [];
                        ns_ind =[];
                        if(State{stage+1}(z,1)==i &&
State{stage+1}(z,2)==0)
                            Temp = 0:i;
                            presched_ind  = Temp';
                            presched_bat =
repmat(State{stage+1}(z,2),[length(Temp) 1]);

                            % Computes the probabilies of "no-show" in
those
                            % transition based prescheduled patints
                            for n=0:length(Temp)-1

ns_ind(n+1,:)=nchoosek(i,n)*(prob_ns_ind^n)*(1-prob_ns_ind)^(i-n);
%binomial distribution for individual
                            end
                            ns_batProb = repmat(1,[length(Temp) 1]); %
probability for the batch
                            ns_indProb = ns_ind;

                            break;
                        elseif( State{stage+1}(z,1)==i &&
State{stage+1}(z,2)==BS)
                            ns_indProb = [];
                            ns_batProb = [];
                            Temp2 = 0:i;

                            for(m=1:length(Temp2))
                                n = m-1;
                                presched_ind = [presched_ind;
repmat(Temp2(m),length(batch),1)];
                                presched_bat = [presched_bat; batch'];
                                ns_ind =
nchoosek(i,n)*(prob_ns_ind^n)*(1-prob_ns_ind)^(i-n); %binomial
distribution for individual
                                ns_indProb = [ns_indProb;
repmat(ns_ind, [length(batch),1])];
                                ns_batProb = [ns_batProb; ns_batch'];
                            end
                            break;
                        end
                    end  % computations done
```

```
                    %                   presched_ind =
max(presched_ind) - presched_ind;  %%Computation of no-shows/// Update
the true prescheduled
                    %                   presched_bat =
max(presched_bat) - presched_bat;
                    % ns_indProb
                    % ns_batProb

                    NumExpStates = length(presched_ind); % number of
possible transitioned states
                    backlog_ind = repmat(
State{stage+1}(z,3),[NumExpStates, 1]);
                    backlog_bat  = repmat( State{stage+1}(z,4),
[NumExpStates, 1]);
                    NoShows = [presched_ind,presched_bat,backlog_ind,
backlog_bat];
                    expStateTemp =  NoShows ; %%%TRANISITION STATES

                    Xk = repmat(State{stage+1}(z,1:2), [NumExpStates,
1]);    %%%original prescheduled
                    Wk = NoShows(:,1:2); % possible no shows
                    expState = Xk - Wk;
                    %                   expStateTemp = repmat([0 0 1
2], [4,1])

                    % Form the possible states the current state can
                    % transition to based on a given decision

                    %                  stageIndx =
findStageIndx(stage);

                    tempExpectation = zeros(1,ukMax+1);
                    if(Decision(1)==0)            %%%%Reject/Don't
Accept

                        Uk{Decision(1)+1} =  [expState, backlog_ind,
backlog_bat];
                        NumPatients = sum(Uk{Decision(1)+1}, 2);
%%%total number of patients in system before any decisions are taken

                        tempExpectation(1) = ComputeProfit(
NumPatients, Uk{Decision(1)+1}, ns_indProb, ns_batProb,
RevenueVal,PenUn, PenOT,...
                            C,N,stage,NumOfStates,  State,
NumExpStates, Expectation,InitialState);

                        %%%Determines who have been serviced prior to
the
                        %%%acceptance of decision (i.e. implied yk=xk-
wk)
                        Overflow{Decision(1)+1} =
ComputeOverFlow(presched_ind,Uk{Decision(1)+1}, StateDim, C);
```

```
                        TrueOverflow = Overflow{Decision(1)+1};
%%%Updates the overflow from original state with the decision
                    end


                    if(Decision(2)==1 && (( State{stage+1}(z,3)+
State{stage+1}(z,4)+ Decision(2))<=MaxB)) %%%%Accept INDIVIDUAL
                        backlog_indUk1 = backlog_ind + Decision(2);
                        Uk{Decision(2)+1} =  [expState,
backlog_indUk1, backlog_bat];


                        tempExpectation(2) =
ComputeProfit(NumPatients,Uk{Decision(2)+1}, ns_indProb, ns_batProb,
RevenueVal,PenUn, PenOT,...
                            C,N,stage,NumOfStates,  State,
NumExpStates, Expectation,InitialState);

                        TrueOverflowUk1 = TrueOverflow;
                        TrueOverflowUk1(:,3) = TrueOverflowUk1(:,3)+
ones(NumExpStates,1);
                        Overflow{Decision(2)+1} = TrueOverflowUk1;
                    else
                        tempExpectation(2) = -inf;

                    end


                    if(Decision(3)== 2 && (( State{stage+1}(z,3)+
State{stage+1}(z,4)+ Decision(3))<=MaxB)) %%%%Accept BATCH (when BS=2)
                        backlog_batUk2= backlog_bat + Decision(3);
                        Uk{Decision(3)+1} =  [expState, backlog_ind,
backlog_batUk2];

                        tempExpectation(3) =
ComputeProfit(NumPatients,Uk{Decision(3)+1}, ns_indProb, ns_batProb,
RevenueVal,PenUn, PenOT,...
                            C,N,stage,NumOfStates,  State,
NumExpStates, Expectation,InitialState);

                        TrueOverflowUk2 = TrueOverflow;
                        TrueOverflowUk2(:,4) = TrueOverflowUk2(:,4)+
2.*ones(NumExpStates,1);
                        Overflow{Decision(3)+1} = TrueOverflowUk2;

                    else
                        tempExpectation(3) = -inf;
                    end
```

```
                    if(Decision(4)== 3 && (( State{stage+1}(z,3)+
State{stage+1}(z,4)+ Decision(4))<=MaxB)) %%%%Accept BATCH (when BS=2)
                        backlog_indUk3 = backlog_ind + Decision(2);
%individual increments by one
                        backlog_batUk3 = backlog_bat + Decision(3);
%batch increments by two
                        Uk{Decision(4)+1} =  [expState,
backlog_indUk3, backlog_batUk3];

                        tempExpectation(4) =
ComputeProfit(NumPatients, Uk{Decision(4)+1}, ns_indProb, ns_batProb,
RevenueVal,PenUn, PenOT,...
                            C,N,stage,NumOfStates,  State,
NumExpStates, Expectation,InitialState);

                        TrueOverflowUk3 = TrueOverflow;
                        TrueOverflowUk3(:,3) = TrueOverflowUk3(:,3) +
ones(NumExpStates,1);
                        TrueOverflowUk3(:,4) = TrueOverflowUk3(:,4) +
2.*ones(NumExpStates,1);
                        Overflow{Decision(4)+1} = TrueOverflowUk3;
                    else
                        tempExpectation(4) = -inf;
                    end

                    %   Compute the optimal expectation
                    [Expectation(z,2*stage+1), index]  =
max(tempExpectation);
                    Expectation(z, 2*stage+2) =  index-1; % optimal
decision

                    OverflowUk = Overflow{index};
                    [ignore WinningOverflowIndex] =
max(sum(OverflowUk, 2));
                    ActualOverflow(z,:) =
OverflowUk(WinningOverflowIndex,:);

                end
            end
            %%%%%%%%%%Displaying Output
            prob_ns_ind2= repmat(prob_ns_ind,[NumOfStates, 1]);
%%%%Used to display probabilities in output
            prob_ns_batch2= repmat(prob_ns_batch,[NumOfStates, 1]);

            export_matrix{count} = [InitialState, State{stage+1},
Expectation, prob_ns_ind2,prob_ns_batch2];
            xlswrite((strcat(strcat(num2str(expr),'TEST_Exp_ALL_'),
date)), export_matrix{count}, count);
        end
    end

 end
```

Compute Overflow

```
function Overflow = ComputeOverFlow(presched_ind,expStateTemp,
StateDim, C)
for j=1:length(presched_ind)
    ykTemp = sum(expStateTemp(j,1:StateDim-1));%%Sum up to individual
backlog
    ykTemp2 = sum(expStateTemp(j,:));%%Sum up to batch backlog
    if sum(expStateTemp(j,1:StateDim-1)) >C      %%Sum up to individual
backlog
        expStateTemp(j,:) = [zeros(1,StateDim-2), ykTemp-C,
expStateTemp(j,StateDim)];
    elseif ykTemp2 <=C
        expStateTemp(j,:) = zeros(1,StateDim);
    elseif ykTemp2 > C      %%Sum up to batch backlog
        expStateTemp(j,:) = [zeros(1,StateDim-1), ykTemp2-C];
    end
end

Overflow = expStateTemp;
```

Compute Expectation

```
function tempExpectation =
ComputeProfit(NumPatients,expState,ns_indProb, ns_batProb,
RevenueVal,PenUn, PenOT , C, N,stage,NumOfStates,...
                           State, NumExpStates,Expectation,
InitialState)

% PenOT=0;
Rev  = RevenueVal.*min(NumPatients , C); %Revenue from serving patients
Unut = PenUn.*max(C-NumPatients, 0); %Unutilized slots
OF   = PenOT.*(max(NumPatients-C, 0)); %Overflow
% OF   = (max(NumPatients-C, 0)+ sum(expState(:,3:4),2)); %Overflow
Profit = Rev-Unut-OF;
% Profit = Rev-Unut;
previousExpCost = zeros( NumExpStates,1); %%%Assuming terminal cost is
ZERO

% RevenueVal=0;

%Calculates the terminal cost for possible backlogs/overflowed patients
if(stage==N-1)
    for b = 1: NumExpStates
        numBacklog= sum(expState(b,3:4));
        if(numBacklog ==0)
            previousExpCost = RevenueVal*numBacklog - zeros(
NumExpStates,1);
        elseif(numBacklog ==1)
            previousExpCost = RevenueVal*numBacklog -(PenOT*ones(
NumExpStates,1));
        elseif(numBacklog ==2)
```

```
            previousExpCost = RevenueVal*numBacklog -(PenOT*2*ones(
NumExpStates,1));
        else
            previousExpCost = RevenueVal*numBacklog -(PenOT*3*ones(
NumExpStates,1));
        end
    end
end

% mapping from previous optimal expected profit(cost)for the particular
state
if(stage~=N-1)
    for k = stage+1: N-1
        for i = 1: NumExpStates
            for j=1:NumOfStates
                if(expState(i,1)==State{k+1}(j,1) &&
expState(i,2)==State{k+1}(j,2)&& expState(i,3)==State{k+1}(j,3)&&
expState(i,4)==State{k+1}(j,4))
                    previousExpCost(i,:) = Expectation(j, 2*(k)+1);
                    %              previousExpCost(i,:) = Expectation(j,
2*(stage-1)+1);
                    break;
                end
            end
            break;
        end
        break;
    end
end

tempExpectation= sum((Profit +
previousExpCost).*ns_indProb.*ns_batProb);
```