June 2024

Aligning Clinical Productivity of Cancer Care Providers: A Simulation-Based Approach

Maile Sinclair-Baxter
University of South Florida

Aligning Clinical Productivity of Cancer Care Providers: A Simulation-Based Approach

by

Maile Sinclair-Baxter

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Science
Department of Industrial and Management Systems Engineering
College of Engineering
University of South Florida

Major Professor: Hadi Gard, Ph.D.
Kingsley Reeves, Ph.D.
Walter Silva Sotillo, Ph.D.

Date of Approval:
June 7, 2024

Keywords: clinical productivity, healthcare providers, data analytics, discrepancy, simulation

Acknowledgments

I would like to acknowledge my major professor, Dr. Hadi Gard. His expertise, guidance, and patience made this project a reality. I would also like to acknowledge Dr. Kingsley Reeves for believing in me from day one, connecting me with the correct resources, and providing assistance with the project.

In addition, I would like to acknowledge Dr. Walter Silva Sotillo for accepting to be part of my committee and for your valuable feedback, and Dr. Grisselle Centeno for listening to my academic goals and connecting me with MyHospital for this amazing project.

Finally, I would also like to thank my husband, Simeon Richard Baxter, and my parents, Marcos Antonio Sinclair Burgess, and Marlene Pons Martinez, for their constant support and prayers.

Table of Contents

List of Tables

List of Figures

Abstract

The growing number of cancer patients over the years has presented many challenges and opportunities for improvement in the healthcare system. These challenges and opportunities have demonstrated the need for a tool that can enhance the clinical productivity of healthcare providers, ultimately impacting patients' overall experience. We studied the clinical Full Time Equivalency (cFTE) for providers at a cancer research center. cFTE is the percentage of time allocated for clinical work. In this cancer research center, the clinical work of each provider is supposed to match a targeted cFTE. The cancer research center's concern was that the provider's clinical productivity was deviating from the targeted cFTE in practice.

We first created a tool to check whether or not there was a discrepancy between the clinical productivity of the providers and the targeted cFTE. This tool revealed that there was, in fact, an average difference of 38.13%. In this study, we developed a simulation-based approach and hypotheses to enhance providers' clinical productivity. By using this simulation tool, we were able to reduce the average difference between the clinical productivity referred to as checked-out cFTE and targeted cFTE to 9.78% and calculate the average difference percentage of improvement between the targeted cFTE and the checked-out cFTE. This value improved to 74.34%.

Chapter 1: Introduction

According to the World Health Organization, “cancer is the second leading cause of death globally, accounting for an estimated 9.6 million deaths, or 1 in 6 deaths, in 2018.” (World Health Organization). Therefore, the need for an established strong system that can manage the number of people who require access to quality cancer care is evident. By doing so, the probability of survival will be higher as people have access to early detection and quality treatment.

As the number of cancer patients and the survival rate of those patients increases, many hospitals are facing a challenge. One area of concern is the disproportionate number of patients assigned to a particular provider in comparison to others. This could be attributed to patient preference or to the referral of cases to providers with particular expertise. Preferences for a provider may be based on their years of experience or the relevant work done in their field. Further, many family care providers refer their patients to colleagues with whom they are familiar or reputable oncologists. Awareness of such preferences and some other constraints, such as physical space, may be leading to disparities, which is important because it affects the provider-to-patient ratio and the quality of care. Measuring clinical productivity can ensure clinical resources are properly allocated to serve patient needs and satisfy a hospital’s mission of quality.

Providers have their time allocated to different responsibilities such as outpatient clinic time, scheduled surgeries, research, and administration time related to clinical duties. During this research, we worked closely with a cancer treatment hospital, which we referred to as MyHospital. This particular cancer center was where the providers of our study performed their clinical duties. The percentage of time allocated for clinical work is measured by clinical Full Time Equivalency (cFTE). cFTE along with individual provider target values to measure clinical productivity, creates an opportunity to have a good appointment scheduling system and patient-to-provider ratio. By analyzing cFTE, the hospital leadership teams can provide the resources needed to manage the number of patients providers are expected to serve.

In MyHospital, three versions of cFTE are considered. The first version is the percentage of time the hospital would like the provider to devote to clinical work. We will refer to this as the targeted cFTE. The second version is the time the provider allocates to clinical work in his or her schedule template that is given to the hospital and used for patient scheduling. We will refer to this as the template cFTE. The third version is the actual percentage of time the provider spent on clinical work as measured using patient data collected as they checked out of the clinic following a patient visit. We will refer to this as the checked-out cFTE. These cFTE measures can vary from one provider to another as the percentage allocated for clinical work may be based on medical specialty and additional responsibilities. For example, we may find a provider with a higher percentage of their time allocated to non-clinical work such as research and teaching. We may also find a provider with a higher percentage of cFTE due to their specialty (e.g., surgery time). These are some of the factors that were considered when conducting an analysis of cFTE.

cFTE is used to measure clinical productivity and capacity and find areas of improvement to better serve patients and providers. The goal of this study was to minimize the discrepancies among the three versions of cFTE: targeted cFTE, template cFTE, and checked-out cFTE for each of the providers analyzed. MyHospital believed that minimizing the discrepancies in cFTE would help decrease instances of under or overutilization of providers.

The targeted cFTE and checked-out cFTE were the parameters for this research. Since the targeted cFTE is given by the hospital to the providers and is part of their contract, and the cFTE checked-out happens after the schedule template is built, the only way to minimize the discrepancy between the targeted cFTE and the checked-out cFTE is through modifications of the template cFTE. By modifying the template and setting it up in a way that meets the targeted cFTE, we expect to see a change in the checked-out cFTE. Template cFTE directly affects the checked-out cFTE as the providers make their template available for appointment scheduling. This research focuses on how to conduct this task. To test this approach, we created a simulation that could simulate the patients coming to the system. This tool simulates the checked-out cFTE. As a result of modifying the template cFTE through strategic adjustments, there was a decrease in the discrepancy between the targeted cFTE and checked-out cFTE by 74.34%.

Chapter 2: Literature Review

In the context of appointment scheduling, there are many factors that affect the utilization of the providers' clinical slots effectively, hence why some providers may not meet their targeted cFTE. The clinical slots are referred to as the slots that providers have available for patients to schedule an appointment. Some of these factors include the number of scheduled clinical slots, patient arrival time, patient wait time, clinical service time, cancellations and no-shows, patient preference, and the provider's available clinic time. Gupta and Denton (2008) described some challenges and opportunities related to appointment scheduling systems in healthcare clinics. The authors mentioned key issues in designing and managing appointment scheduling systems in their paper. Some issues discussed were the patient arrival process, clinical service process, patient and provider preferences, incentives, and performance measures. In our paper, we focused on relevant literature pertaining to two categories: attendance management challenges and service forecasting. There is an extensive amount of related work that has reviewed attendance management challenges and forecasting. In this section, we will provide more details about each of these areas of research. The first topic of the literature review, attendance management challenges, includes overbooking or double booking to minimize the costs caused by no-shows. The second part of the literature review will be focused on forecasting. Using different prediction models to forecast the probability of no-shows can benefit all medical facilities by having a more structured system. A more structured system can be achieved by having a better scheduling appointment. This will allow patients to have access to health care and improve the experience for both the patients and the medical providers.

2.1 Attendance Management Challenges

No-shows and cancellations have been studied in many papers. Past studies have examined some patterns and factors that can cause no-shows, such as logistical challenges, forgetting about the appointment, socioeconomic status, age, health conditions, and emotions such as fear. The data presented by these studies were collected through various methods such as patient interviews (Lacy et al., 2004), or patient records (Davies et al., 2016) that identified patterns such as gender, males had a higher no-show rate than females, or younger patients or new patients were more prone to miss their appointments. Coppa (2023) developed a prediction for non-arrival (no-shows) to scheduled ambulatory appointments. One of the key results showed that patients with appointments that were no-shows were more likely to be single, racial/ethnic minority, and not have an established primary care provider compared to those who arrived at their appointment.

Adjustments such as overbooking have been suggested to counteract the impact of no-shows on the appointment scheduling. Nasir (2020) presented a data analytics framework to study the factors of no-shows through machine learning models with the purpose of predicting whether a patient would be a no-show. Their study also proposed a methodology that would integrate a prediction model with the purpose of creating an overbooking decision tool with variable overbooking times in different time slots. The prediction model was intergraded using Bayesian inference system. The study showed that it is possible to identify a patient at higher risk of not attending their appointment, and intervention techniques can be used by clinics, such as call or text reminders.

In MyHospital, patients are allocated to a type of appointment slot and duration depending on if they are new patients or established patients. New patients may have a longer appointment time slot than established patients. To prevent unexpected delays and extend service duration, the doctors have assistant physician assistants, fellows, or nurses who start the consultation with the patients if there is an overbooking or two appointments scheduled at the same time. Zacharias and Pinedo (2014) developed strategies and mathematical models that minimized the impact of no-shows without compromising the quality of service. A strategy mentioned in the paper was the use of an overbooking model for scheduling arrivals at a medical facility under a no-show environment. Their proposed heuristic approach performed better as the variability in the no-show probabilities increased.

LaGanga and Lawrence (2012) also studied the idea of overbooking. This involved scheduling more patient appointments than the available time slots. The authors' study proposed a gradient search solution algorithm, which added new appointments in order of maximum contribution to utility. In a different study, LaGanga and Lawerence (2007) developed a utility function to achieve a trade-off between the benefits and risks of overbooking. The results of the study showed the benefits of overbooking by compressing the appointment interval if the clinic or hospital experienced high demand, low service time variability, and high no-show rates. In that same study, the authors also conducted a simulation study that concluded appointment shortening was a method that could be used to reduce patient waiting time, while double booking could reduce provider overtime.

Liu and Ziya (2014) introduced mathematical models such as stylized single server queueing models, which were considered to model the appointment scheduled for a provider. They presented two scenarios. The first scenario presented the service capacity, such as the overbooking level, as fixed, and the decision variable was the panel size (number of patients under the care of a provider). The second scenario had both the panel size and the service capacity as the decision variables. The study revealed how demand and capacity decisions could be adjusted in response to patients' no-show behavior due to patients' sensitivity to incremental delays. This scenario may also help forecast the number of patients that would show up to their allocated appointment slots based on behavioral analysis. Srinivas and Ravindran (2018) proposed a prescriptive analytics framework where patient-specific no-show risk predicted using multiple data sources were utilized to develop dynamic scheduling rules to improve the performance of patient appointment scheduling. This was concerning patient satisfaction and resource utilization. The dynamic scheduling rules were a combination of two factors: sequencing policy and overbooking policy. The overbooking policy identified the type of patients who could be scheduled together in the same slot. The evaluation of the scheduling rules based on the two factors indicated that integrating the risk of patient no-shows with patient scheduling led to better operational performance in comparison to rules that randomly overbooked patients with the same expected no-show rate. In addition, scheduling patients based on expected service times and overbooking a combination of low-risk and high-risk patients in the first available slot resulted in the minimum total cost. These optimization strategies, combined with the implementation of machine learning, seek to improve patient access and service, healthcare resources allocation, reduce wait times, and possibly reduce the number of no-shows or cancellations.

Muthuraman and Lawley (2008) developed a stochastic overbooking model and a myopic scheduling policy for outpatient clinics by categorizing patients into groups based on their no-show probability estimates. The scheduler assigned patients to slots sequentially through a patient call-in process, where the scheduler provided each calling patient with an appointment when the call arrived. The scheduling objective function captured patient waiting time, staff overtime, and patient revenue.

Klassen and Yoogalingam (2009) proposed a simulation-optimization approach to determine the optimal rules for a stochastic appointment scheduling problem for outpatient healthcare services. Some of the major findings were that the plateau-dome scheduling pattern was robust across different factors (performance measures, number of appointments, appointment lengths, cost factors, levels of no-shows, and session lengths, among others) and was adjusted depending on the environment. Some other findings mentioned in the paper were the presence of no-shows introducing the strategy of double-booking. In this section, the results of the simulation suggested that if overbooking was not used, a good alternative was to book the first two patients at the beginning of the session to shorten the appointment slots on the plateau. A restrictive session end time for the provider resulted in client waiting, reducing the performance of the appointment system.

2.2 Forecasting

Another line of research focuses on different statistical methods used to conduct predictive analysis of different appointment scheduling system problems and find practical solutions for them. One of the works that study the method of forecasting in appointment scheduling is Robinson and Chen (2011), who introduced a queuing-based approach for estimating the server's perceived relative cost of customer waiting based on the observed average number of customers waiting and the server's utilization. Samorani and Laganga (2015) proposed a combination of predictive analytics, optimization, and overbooking to maximize the number of outpatient appointments seen without increasing waiting time and provider overtime. Individual no-show predictions were used to accomplish this based on individual appointment characteristics and appointment day. A column-generation solution method was developed to schedule appointments given individual day-dependent show predictions. Based on the decision analysis made by the column-generation method, the authors developed a policy that consisted of scheduling the predicted shows in S-slots in the near future and the predicted no-shows in the N-slots farther in the future. As one of their study observations, with this model, if the patient does not show, their appointment slot will be partly used by appointments assigned to the preceding S-slots.

The authors (Robaina, Bastrom, Richardson, & Edmonds, 2020) evaluated the no-show rate in the pediatric orthopedic population via electronic health record (HER). They identified potential factors predictive of a higher no-show rate. Classification and regression trees (CART) were constructed to identify predictors of no-shows. The data included appointments, and different variables were taken into consideration, such as appointment status, age, gender, type of visit, payor type (government vs private insurance), distance of residence to clinic, region of residence, clinic location, clinic type, and appointment day of the week, hour, and month. The results demonstrated that factors such as payor type and longer duration between scheduling and appointment played a role in predicting non-attendance at outpatient orthopedic appointments. Clinics can develop intervention practices to target these specific areas.

On the other hand, Salah and Sriniva (2022) introduced the idea of a predict-then-schedule framework, which was proposed for the design of an appointment system for sequentially scheduling patients in the presence of factors such as no-show and service time uncertainty, and two patient classes. In the predict step, the patient no-show risk and service duration were estimated using a machine-learning model. The schedule step calculated the patient's appointment time and interval by integrating the predictions with three scheduling decisions: allocation, sequencing, and overbooking. The study demonstrated the capability of Machine Learning (ML) algorithms to accurately predict patient service time and no-shows, which ultimately showed improvement in appointment system efficiency for all the clinical environments tested in a sequential scheduling framework. Grant, Gurvich, Mutharasan, and Van Mieghem (2022) studied the risk of incurring costly failures such as readmissions in healthcare or engine failures when appointments are delayed in dynamic stochastic appointment scheduling. This study was based on a preventative maintenance setting. The authors analyzed the use of surge capacity by scheduling an appointment sooner versus delaying it and facing the failure cost. The stochastic dynamic programming (DP) formulation was used, and the results showed some insights on how to schedule preventative appointments and use surge capacity for the costs of appointments and failures.

To address the negative impact that no-shows have on clinics, Harris and Samorani (2021) studied predictive overbooking systems. The study focused on how to select the no-show prediction model that had the greater classification performance. The probability of appointment requests that would be a no-show was predicted by a probabilistic classifier. A scheduling algorithm used individual predictions to schedule appointments optimally. They found that the cost of a schedule when Brier score or Log loss is used to choose a classifier was 4.5% less than the cost of a schedule when area under the receiver operator characteristic curve (AUC) was used to choose a classifier for models that were solved without heuristic and 7% for larger scheduling models with a heuristic.

The hybrid data mining simulation methodology used by Glowacka, Henry, May (2009) highlighted the importance of identifying differences in no-show rates and how the ability to integrate variable no-show rates, instead of treating the probability of not showing up for an appointment as a constant value, in this case, obtained using the association rule mining (ARM) approach, into simulation models can improve the performance of complex scheduling systems in outpatient clinics.

Huang and Zuniga (2012) showed that overbooking policies differed from physicians, clinics, and specialties. The authors believed that the most appropriate overbooking approach is one that accounts for individual patient's conditions, and the prediction of the probability of a patient's no-show is high regardless of the type of specialty. For the approach, the authors proposed that each clinic determined the least amount of patients to schedule without overbooking and then used their proposed overbooking system to find where and how many appointments to overbook without increasing patient wait time, provider idle time, and overtime, ultimately minimizing the total cost.

Our review of the literature indicates that many studies focus on overbooking or double booking as a method to minimize the impact of no-shows on appointment system scheduling or to care for more patients. In MyHospital, some providers have already implemented overbooking. In our study, we consider other factors besides no-shows, such as administrative tasks and the structure of the schedule, as possible causes of the schedule not being optimally utilized and at capacity. We do need to take into consideration that the solution we propose does not cause overtime. These findings are particularly relevant for the healthcare sector as they provide a comprehensive understanding of the various factors that can affect appointment scheduling systems and offer strategies for optimization.

Chapter 3: Problem Description and Hypotheses

MyHospital is a cancer treatment and research center with the mission to provide diagnosis and treatment plans to its patients. In addition to providing cancer care, MyHospital has 16 providers in the outpatient center that is the focus of our study. In this study, we analyzed each of those providers. These providers have different clinical responsibilities besides seeing patients; some of those responsibilities are determined by the provider’s specialty. The providers’ clinical responsibilities are measured by the clinical Full Time Equivalency (cFTE), which is defined as the percentage of time allocated for clinical work. Each provider has a targeted cFTE, a template cFTE, and a checked-out cFTE. The targeted cFTE is the percentage of time that the hospital would like the provider to devote to clinical work. The template cFTE is the time that the provider allocates to clinical work in his or her schedule template that is given to the hospital and used for patient scheduling. The checked-out cFTE is the actual percentage of time that the provider spent on clinical work as measured using patient data collected as they checked out of the clinic following a patient visit. MyHospital is concerned with the providers meeting the targeted cFTE. Providers are supposed to build their schedule or template based on the targeted cFTE. Thus, we focused on analyzing the clinical responsibilities of the 16 providers.

While targeted cFTE is provided by MyHospital, calculating template and checked-out cFTE required a specific tool, which will be addressed in Chapter 4. Based on the data, the concern of MyHospital is to correct the discrepancy between the targeted cFTE and the checked-out cFTE. We developed a tool that allowed us to calculate the average difference between the two cFTEs. There was a 38.13% average difference. This average difference showed the large deviation of the checked-out cFTE from the targeted cFTE with the targeted cFTE being the largest between the two values. This could suggest that the providers' templates were not built to have as many clinical slots for patients available as the targeted cFTE intended. Our goal of the research is to address MyHospital's concerns by minimizing this discrepancy between targeted cFTE and checked-out cFTE. Since we have no direct control over the targeted cFTE and checked-out cFTE, our approach to reduce this discrepancy is to manipulate the template cFTE.

Discrepancies between the targeted cFTE and the checked-out cFTE may be attributed to over-bookings, overlapping appointments, missed appointments, not enough support available to the providers to meet the demand, and discontinuation of provider contracts, among other reasons. Additionally, a provider may later discover that there are more available appointment slots than anticipated for patient visits. This may indicate that the provider's schedule is not made to meet the targeted cFTE as it does not show the capacity that they have to meet with more patients. The overbookings are a result of appointments that are added after the schedule has been released and they are not slotted. In MyHospital, it is understood that providers only schedule overlapping appointments when they have the support of Advanced Practice Providers (APPs) or a fellow (a physician in training for a subspecialty).

We had some raw data provided by MyHospital for the template and checked-out data for each provider. The template data had information such as the clinical slots, the dates, the start and the end time, the type of patient, and double booking, among other useful information. The checked-out data showed what was actually happening in the system. The data had the dates, check-in time as well as the checked-out time, no-shows, double booking, etc. These data sets were important to calculate the template cFTE and the checked-out cFTE. As part of our initial analysis of cFTE, we collaborated with our partners at MyHospital to obtain more information on how they calculated the cFTE. We needed to develop a tool, which will be explained in Chapter 4, to calculate the template cFTE and the checked-out cFTE for both data sets for each provider to check if the results were going to be the same as the targeted cFTE. We validated that the tool worked as we were able to calculate the cFTEs. Our calculations proved, however, that there was, in fact, a discrepancy between the targeted cFTE checked-out cFTE. A histogram was created to show the number of patients, the type of patients seen by the providers, and the number of clinical days that each provider had was conducted based on the template cFTE data to determine the factors causing the discrepancies. From the statistical analysis, we created a list of existing principles that were found in the template cFTE data. Based on the existing principles in the template and suggested principles from the checked-out cFTE, we created a list of suggested hypotheses to test:

- By increasing the number of clinical days for certain providers in the template cFTE provided by MyHospital, we expect to observe an increase in the checked-out cFTE.
- By increasing the number of clinical slots for appointments in the morning, we expect to observe an increase in the checked-out cFTE.
- By increasing the number of clinical slots for appointments in the afternoon, we expect to observe an increase in the checked-out cFTE.

We formally defined the hypotheses in Chapter 4. With these hypotheses, our goal is not to develop a tool to generate templates for the providers but to adjust the templates provided by MyHospital.

Chapter 4: Methodology

The typical Full-Time Equivalent (FTE) is calculated by dividing the employee's scheduled hours by their official workweek schedule. As previously mentioned in Chapter 3, we created some hypotheses to test, expecting a decrease in the discrepancy between checked-out cFTE and targeted cFTE. The hypotheses consist of the following:

- Hypothesis I: By increasing the number of clinical slots in the morning for a provider with enough clinical slots in the afternoon, we expect to observe an increase in the checked-out cFTE.
- Hypothesis II: By increasing the number of clinical slots in the afternoon for a provider with enough clinical slots in the morning, we expect to observe an increase in the checked-out cFTE.
- Hypothesis III: By increasing the number of clinical days for a provider that has an adequate number of clinical slots in the mornings and the afternoons on their existing clinical days but checked-out cFTE is still lower than the targeted cFTE, we expect to observe an increase in the checked-out cFTE.

These hypotheses were created based on observations made from analysis conducted on the raw template data for each provider. To test our hypotheses, we needed two tools. One tool was a data preparation tool to transform raw input data into a format conducive to computing the template cFTE and the checked-out cFTE values. The second tool was a simulation tool to replicate the conditions of our hypotheses.

4.1 Data Preparation Tool

We were provided with raw data for template cFTE and checked-out cFTE. The template data provides the scheduled time that the provider allocates to clinical work. On the other hand, the checked-out data reflects the actual time that the provider spent on clinical work as measured using patient data collected as they checked out of the clinic. The template data contained the start date, end date, start time, and end time for each appointment slot. Additionally, the appointment duration was also included, as this could differ from one patient to another, depending on appointment type and patient (i.e., new patient or established patient). The checkED-out data contained the same information as the template data in addition to the time of patient check-in, their checked-out, and whether the appointment was an overbook. Even though the data was useful, we needed to develop a tool that made the data feasible to calculate the template cFTE and checked-out cFTE for each provider.

For the initial analysis, we wanted to capture and report the number of patients entering and leaving the system and the number of patients in the system at each given time. This information is required to calculate the cFTE for both template and checked-out. In order to obtain this information, we used two programming languages, R and Matlab (Algorithm 1 and Algorithm 2). The output using the R code contained three columns. Column A, labeled "timestamp", the timestamp or the time intervals based on the provider's data. In Algorithm 1, lines 13 through 15, the code creates a time sequence with the start date and end date of our data based on five-minute intervals. This is how the timestamp column was created. Column B, labeled "counter", shows when a patient enters the system, represented with +1, and when a patient is leaving the system, represented with -1. Lines 1 through line 10 create a counter variable using the counter function that captures every time a patient enters and leaves the system. Column C contains the number of patients in the system at a specific time, labeled “volume”. The final version of the data, which was the one used to proceed further with the initial analysis, was obtained using Matlab. Algorithm 2 presents a loop that was created, which checks if the numbers contained in column B are greater than zero. If the condition is true, those numbers are copied in column D; otherwise, they are copied in column E. The data will contain five columns. The first three columns are the same as previously described with the data captured by the code created in R. The two additional columns of the final version contain the same information as column B but are separated. Column D represents the patients when the patients check-in in the system, labeled “In”. Column E represents the patients that leave the system, labeled “Out” (Figure 4.3 and 4.4).

Algorithm 1 Counting Procedure in R

```
##count if the previous appointment is the same as the previous one in Template
1. P9_T<-P9_T %>%
2.  mutate
(STDT=floor_date(STDT,unit='minutes'),ETED=floor_date(ETED,unit='minute'))

3. APPSTD9T<-P9_T %>%
4.     select(timestamp=STDT) %>%
5.     mutate(counter=1)

6. APPETD9T<-P9_T %>%
7.       select(timestamp=ETED) %>%
8.       mutate(counter=-1)

 9.  template_volumes_9<-APPSTD9T %>%
10.    bind_rows(APPETD9T) %>%
11.    arrange(timestamp,counter) %>%
12.    mutate(volume_9=cumsum(counter))

#Creating a sequence of times from start to end
13. P9_start<-min(template_volumes_9$timestamp)
14. P9_end<-max(template_volumes_9$timestamp)
15. time_period_9<-tibble(timestamp=seq(P9_start,P9_end,by='5 mins'))

16. template_volumes_9<-template_volumes_9 %>%
17.    right_join(time_period_9, by='timestamp') %>%
18.    arrange(timestamp) %>%
19.    fill(volume_9, .direction = 'down')
20. template_volumes_9
```

Figure 4.1 Algorithm 1 Counting Procedure in R.

Algorithm 2 Matlab Code to Remove the Repeated Dates

```
[m,n]= size(Data_7);
if n<=3
  Data_7 = [Data_7 cell(m,2)];
end
Data_7{1,4} = 'In';
Data_7{1,5} = 'Out';
ii = 2;
while true
   if Data_7{ii,2} >0 && (isempty(Data_7{ii,4}) || isnan(Data_7{ii,4}))
      Data_7{ii,4} = Data_7{ii,2};
   elseif Data_7{ii,2} <0 && (isempty(Data_7{ii,5}) || isnan(Data_7{ii,5}) )
        Data_7{ii,5} = -1*Data_7{ii,2};
   end
   if ii>=length(Data_7(:,1))
       break
   end
   if length(Data_7{ii,1})==length(Data_7{ii+1,1})&& all(Data_7{ii,1}==Data_7{ii+1,1})
       if ~isnan(Data_7{ii+1,2})
           Data_7{ii,3} = Data_7{ii,3} + Data_7{ii+1,2};
           Data_7{ii,2} = Data_7{ii,2} + Data_7{ii+1,2};
             if Data_7{ii+1,2} >0
               if isempty(Data_7{ii,4}) || isnan(Data_7{ii,4})
                  Data_7{ii,4} = Data_7{ii+1,2};
               else
                   Data_7{ii,4} = Data_7{ii,4}+Data_7{ii+1,2};
               end
           elseif Data_7{ii+1,2} <0
               if isempty(Data_7{ii,5}) || isnan(Data_7{ii,5})
                  Data_7{ii,5} = -1*Data_7{ii+1,2};
               else
                  Data_7{ii,5} = Data_7{ii,5}-1*Data_7{ii+1,2};
               end
           end
      end
      Data_7(ii+1,:) = [];
  else
      ii = ii+1 ;
  end
end
```

Figure 4.2 Algorithm 2 Matlab Code to Remove the Repeated Dates.

	A	B	C	D	E
1	timestamp	counter	volume	In	Out
2	7/5/2018 8:00	1	1	1	
3	7/5/2018 8:15		1		
4	7/5/2018 8:30		1		
5	7/5/2018 8:45		1		
6	7/5/2018 9:00	-1	0		1
7	7/5/2018 9:15		0		
8	7/5/2018 9:30	1	1	1	
9	7/5/2018 9:45		1		
10	7/5/2018 10:00	0	1	1	1
11	7/5/2018 10:15		1		
12	7/5/2018 10:30	0	1	1	1
13	7/5/2018 10:45		1		
14	7/5/2018 11:00	0	1	1	1
15	7/5/2018 11:15		1		
16	7/5/2018 11:30	0	1	1	1
17	7/5/2018 11:45		1		
18	7/5/2018 12:00	-1	0		1
19	7/5/2018 12:15		0		

Figure 4.3 Provider 3 Template Output Data.

	A	B	C	D	E
1	timestamp	counter	volume	In	Out
2	7/5/2018 8:00	1	1	1	
3	7/5/2018 8:15		1		
4	7/5/2018 8:30		1		
5	7/5/2018 8:45		1		
6	7/5/2018 9:00	0	1	1	1
7	7/5/2018 9:15	1	2	1	
8	7/5/2018 9:30	0	2	1	1
9	7/5/2018 9:45	-1	1		1
10	7/5/2018 10:00	0	1	1	1
11	7/5/2018 10:15		1		
12	7/5/2018 10:30	0	1	1	1
13	7/5/2018 10:45		1		
14	7/5/2018 11:00	0	1	1	1
15	7/5/2018 11:15		1		
16	7/5/2018 11:30	1	2	2	1
17	7/5/2018 11:45		2		
18	7/5/2018 12:00	-1	1	1	2
19	7/5/2018 12:15		1		
20	7/5/2018 12:30	0	1	1	1
21	7/5/2018 12:45		1		
22	7/5/2018 13:00	0	1	1	1
23	7/5/2018 13:15		1		
24	7/5/2018 13:30	-1	0		1

Figure 4.4 Provider 3 Checked-Out Output Data.

When calculating the cFTEs, noticeable differences were observed between the value calculated using the template data and the checked-out data, as tabulated in Table 4.2. The template cFTE and checked-out cFTE percentages were expected to be similar, as they both should reflect the percentage of the time the provider allocated to seeing patients.

Some acronyms used in Table 4.2 are defined in Table 4.1.

Table 4.1 Acronyms

Acronym	**Definition**
Provider #	Provider Number
TAcFTE	Targeted cFTE
TEcFTE	Template cFTE
COcFTE	Checked-Out cFTE

Table 4.2 Targeted cFTE, Template cFTE and Checked-out cFTE Calculations

Provider #	**TAcFTE**	**TEcFTE**	**COcFTE**
1	0.10	0.10	0.09
2	0.12	0.09	0.10
3	0.06	0.11	0.07
4	0.13	0.10	0.12
5	0.36	0.21	0.24
6	0.45	0.13	0.11
7	0.50	0.34	0.29
8	0.52	0.41	0.37
9	0.80	0.46	0.41
10	0.43	0.24	0.24
11	0.42	0.38	0.38
12	0.36	0.26	0.25
13	0.39	0.28	0.26
14	0.80	0.36	0.29
15	0.40	0.06	0.02
16	0.39	0.15	0.17

4.2 Simulation Tool

The second tool used to test our hypotheses was the recreation of a simulation tool that could replicate the conditions of our hypotheses. In order to test our hypotheses, we needed to modify the template. To see how that performed in practice, we needed a simulator to simulate how the patients were entering the system. Using this simulation, we collected the data needed to calculate the checked-out cFTE.

In MyHospital, patients are categorized as established patients, new patients, or others. The category “others” includes patients that do not fall under the established patient or the new patient category. The simulation system was intended to generate these patients, identify the type of patient, and identify whether or not the patients were going to show up in the system. To accomplish this, we needed to calculate the probability of a patient entering the system for that time slot and the probability of a no-show. This probability would also tell us whether the patient would be an established patient, a new patient, or other. These probabilities were collected from the existing data set previously provided by MyHospital; particularly, we collected the total number of observed clinical slots, the total number of established patients, new patients, or others, using a PivotTable. On average, we had access to one year of data, and each provider had a total of 616 established patients, 133 new patients, 126 categorized as others, and 22 no-shows. Table 4.3 shows the probabilities calculated for provider 1.

As part of the analysis, we wanted to validate the results calculated for the template cFTE and the checked-out cFTE. This simulation was supposed to give us the same cFTE result as the ones tabulated in Table 4.2. We ran the simulation for each of the providers and obtained the same results for the template cFTEs and the checked-out cFTEs previously calculated in section 4.1. Figure 4.5 is a representation of part of the simulation performed in Excel for one of the providers.

Table 4.3 Probability Calculation for Provider 1

	Established Patient	**New Patient**	**Others**
Probability of Patients	0.89	0.08	0.04
Probability of No-Shows	0.01	0	0

	A	B	C	D	E	F	G	H	I	J	K
1		EP	NP	Others							
2	Probability (patients)	0.66	0.0175	0.3225							
3	Probability for no shows	0.041667	0	0.054264							
4											
5	timestamp	counter	volume_9	Assumptic	Rand	Label	Rand	No Show			cFTE
6	2018/07/03 09:00:00	1	1		0.894026	Others	0.870973	1	5	15595	0.128927
7	2018/07/03 09:05:00		1		0.957174	Others	0.730663	1	5		
8	2018/07/03 09:10:00		1		0.92318	Others	0.01633	0	0		
9	2018/07/03 09:15:00		1		0.247348	EP	0.56661	1	5		
10	2018/07/03 09:20:00		1		0.635919	EP	0.419253	1	5		
11	2018/07/03 09:25:00		1		0.237295	EP	0.274996	1	5		
12	2018/07/03 09:30:00	1	2		0.086773	EP	0.612166	1	5		
13	2018/07/03 09:35:00		2		0.530529	EP	0.322044	1	5		
14	2018/07/03 09:40:00		2		0.855075	Others	0.635277	1	5		
15	2018/07/03 09:45:00		2		0.497832	EP	0.969591	1	5		
16	2018/07/03 09:50:00		2		0.987886	Others	0.680959	1	5		
17	2018/07/03 09:55:00		2		0.884801	Others	0.068386	1	5		
18	2018/07/03 10:00:00	-1	1		0.874656	Others	0.308565	1	5		
19	2018/07/03 10:00:00	-1	0		0.492724	-	0.16893	0	0		
20	2018/07/03 10:00:00	1	1		0.411719	EP	0.139918	1	5		
21	2018/07/03 10:05:00		1		0.606169	EP	0.124488	1	5		
22	2018/07/03 10:10:00		1		0.85146	Others	0.162069	1	5		
23	2018/07/03 10:15:00		1		0.438499	EP	0.190682	1	5		
24	2018/07/03 10:20:00		1		0.935653	Others	0.30976	1	5		
25	2018/07/03 10:25:00		1		0.164036	EP	0.11135	1	5		
26	2018/07/03 10:30:00	-1	0		0.202033	-	0.259436	0	0		
27	2018/07/03 10:30:00	1	1		0.568773	EP	0.112956	1	5		
28	2018/07/03 10:35:00		1		0.754517	Others	0.624649	1	5		
29	2018/07/03 10:40:00		1		0.622652	EP	0.892515	1	5		
30	2018/07/03 10:45:00		1		0.621153	EP	0.233512	1	5		
31	2018/07/03 10:50:00		1		0.578525	EP	0.710205	1	5		
32	2018/07/03 10:55:00		1		0.250647	EP	0.012562	0	0		
33	2018/07/03 11:00:00	-1	0		0.196054	-	0.836127	0	0		
34	2018/07/03 11:00:00	1	1		0.074755	EP	0.004876	0	0		
35	2018/07/03 11:05:00		1		0.979357	Others	0.179839	1	5		
36	2018/07/03 11:10:00		1		0.530698	EP	0.012322	0	0		

Template

Figure 4.5 Simulation.

Once the validation part of the analysis was completed, we used the simulation to test our hypotheses. The result of this simulation is discussed in Chapter 5.

Chapter 5: Results and Future Research Directions

5.1 Results

As previously mentioned, we created some hypotheses to test, expecting a decrease in the discrepancy between checked-out cFTE and targeted cFTE. The hypotheses applied consist of the following:

- Hypothesis I: By increasing the number of clinical slots in the morning for a provider with enough clinical slots in the afternoon, we expect to observe an increase in the checked-out cFTE.
- Hypothesis II: By increasing the number of clinical slots in the afternoon for a provider with enough clinical slots in the morning, we expect to observe an increase in the checked-out cFTE.
- Hypothesis III: By increasing the number of clinical days for a provider that has an adequate number of clinical slots in the mornings and the afternoons on their existing clinical days but checked-out cFTE is still lower than the targeted cFTE, we expect to observe an increase in the checked-out cFTE.

The average difference between checked-out and target cFTE before the hypotheses were applied was 38.13%. Once the hypotheses were applied, the average difference decreased to 9.78%. We originally had 16 providers, but two were removed for this part of the simulation. As observed in Table 5.3, there were two providers that we could not properly conduct the analysis because they had a very small number of data points; thus, these two providers were removed from the simulation analysis. The four providers (2,4,11,14) had a more complex schedule as they were surgeons. Surgeons have a different type of template due to operation room (O.R) constraints. For these four providers, we decided not to make any further changes due to not having more information about the constraints of the surgeons. The hypotheses that were applied to each provider are shown in Table 5.1.

Table 5.1 Hypotheses Applied

# Provider	# Hypothesis Applied
1	I and II.
2	No hypotheses were applied since there was no discrepancy between their cFTEs.
3	II.
4	Surgeon- No hypotheses applied.
5	I and III.
6	III.
7	III.
8	I and II.
9	III.
10	III.
11	Surgeon- No hypotheses applied.
12	III.
13	I and II.
14	Surgeon- No hypotheses applied.

Table 5.2 Acronyms

Acronym	Definition
Provider #	Provider Number.
CO-cFTE	Checked-Out cFTE.
Old	Original Setting.
New	After applying hypotheses and running simulation.
Discrepancy %	Average Percentage Difference (Checked-out cFTE and Targeted cFTE) %.
NA	No Principles Added.

Table 5.3 Results of the Simulation

Provider #	Target cFTE	Old		New	
		CO-cFTE	Discrepancy %	CO-cFTE	Discrepancy %
1	0.1	0.09	13.00	0.1	0
2	0.06	0.06	10.00	0.06	NA
3	0.13	0.12	7.69	0.13	0
4	0.36	0.24	33.33	NA	NA
5	0.45	0.11	75.56	0.4	11.11
6	0.50	0.29	42.00	0.41	18.00
7	0.52	0.37	28.85	0.53	1.92
8	0.80	0.41	48.75	0.64	20.00
9	0.43	0.24	44.19	0.4	6.98
10	0.42	0.38	9.52	0.44	4.76
11	0.36	0.25	30.56	NA	NA
12	0.39	0.26	33.33	0.38	2.56
13	0.80	0.29	63.75	0.54	32.50
14	0.39	0.17	56.41	NA	NA

Our goal for this research was to reduce the discrepancy between the targeted cFTE and the checked-out cFTE by manipulating the template cFTE. We accomplished this by achieving an average percentage of improvement difference of 74.34%. Based on these results, we may say that some adjustments to the template built by the providers are necessary to accommodate more patients during assigned clinical days and, hence, meet the targeted cFTE. These results may also provide more information to MyHospital about provider utilization and capacity. In addition, this may also lead to projects that can provide and help allocate resources to support the providers and meet the hospital's demand. As a result of these implementations, more patients may be served and have access to quality healthcare.

5.2 Future Research Directions

For future research directions, we will examine more closely the development of a tool for directly creating template cFTE rather than treating it as a black box and solely manipulating its outcome. Further analysis and exploration will be needed to decrease the average difference percentage between the targeted cFTE and the checked-out cFTE or to eliminate the discrepancy completely between the two without affecting the providers' other administrative responsibilities.

References

Coppa, K., Kim, E. J., Oppenheim, M. I., Bock, K. R., Zanos, T. P., & Hirsch, J. S. (2023). Application of a Machine Learning Algorithm to Develop and Validate a Prediction Model for Ambulatory Non-Arrivals. *Journal of General Internal Medicine*, *38*(10), 2298–2307. https://doi.org/10.1007/s11606-023-08065-y

Davies, M. L., Goffman, R. M., May, J. H., Monte, R. J., Rodriguez, K. L., Tjader, Y. C., & Vargas, D. L. (2016). Large-scale no-show patterns and distributions for clinic operational research. *Healthcare (Switzerland)*, *4*(1). https://doi.org/10.3390/healthcare4010015

Glowacka, K. J., Henry, R. M., & May, J. H. (2009). A hybrid data mining/simulation approach for modelling outpatient no-shows in clinic scheduling. *Journal of the Operational Research Society*, *60*(8), 1056–1068. https://doi.org/10.1057/jors.2008.177

Grant, B., Gurvich, I., Mutharasan, R. K., & van Mieghem, J. A. (2022). Optimal Dynamic Appointment Scheduling of Base and Surge Capacity. *Manufacturing and Service Operations Management*, *24*(1), 59–76. https://doi.org/10.1287/MSOM.2020.0932

Gupta, D., & Denton, B. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions (Institute of Industrial Engineers)*, *40*(9), 800–819. https://doi.org/10.1080/07408170802165880

Harris, S. L., & Samorani, M. (2021). On selecting a probabilistic classifier for appointment no-show prediction. *Decision Support Systems*, 142. https://doi.org/10.1016/j.dss.2020.113472

Huang, Y., & Zuniga, P. (2012). Dynamic overbooking scheduling system to improve patient access. *Journal of the Operational Research Society*, *63*(6), 810–820. https://doi.org/10.1057/jors.2011.83

Klassen, K. J., & Yoogalingam, R. (n.d.). Improving Performance in Outpatient Appointment Services with a Simulation Optimization Approach. https://doi.org/10.3401/poms.1080.01021

Lacy, N. L., Paulman, A., Reuter, M. D., & Lovejoy, B. (2004). Why we don't come: Patient perceptions on no-shows. *Annals of Family Medicine*, *2*(6), 541–545. https://doi.org/10.1370/afm.123

Muthuraman, K., & Lawley, M. (2008). A stochastic overbooking model for outpatient clinical scheduling with no-shows. IIE Transactions (Institute of Industrial Engineers), 40(9), 820–837. https://doi.org/10.1080/07408170802165823

Laganga, L. R., & Lawrence, S. R. (2007). Clinic overbooking to improve patient access and increase provider productivity. *Decision Sciences*, *38*(2), 251–276. https://doi.org/10.1111/j.1540-5915.2007.00158.x

Laganga, L. R., & Lawrence, S. R. (2012). Appointment overbooking in health care clinics to improve patient service and clinic performance. *Production and Operations Management*, *21*(5), 874–888. https://doi.org/10.1111/j.1937-5956.2011.01308.x

Liu, N., & Ziya, S. (2014). Panel size and overbooking decisions for appointment-based services under patient no-shows. *Production and Operations Management*, *23*(12), 2209–2223. https://doi.org/10.1111/poms.12200

Nasir, M., Summerfield, N., Dag, A., & Oztekin, A. (2020). A service analytic approach to studying patient no-shows. *Service Business*, *14*(2), 287–313. https://doi.org/10.1007/s11628-020-00415-8

Robaina, J. A., Bastrom, T. P., Richardson, A. C., & Edmonds, E. W. (2020). Predicting no-shows in paediatric orthopaedic clinics. *BMJ Health and Care Informatics*, *27*(1). https://doi.org/10.1136/bmjhci-2019-100047

Robinson, L. W., & Chen, R. R. (2011). Estimating the implied value of the customer's waiting time. *Manufacturing and Service Operations Management*, *13*(1). https://doi.org/10.1287/msom.1100.0304

Salah, H., & Srinivas, S. (2022). Explainable machine learning framework for predicting long-term cardiovascular disease risk among adolescents. *Scientific Reports*, *12*(1). https://doi.org/10.1038/s41598-022-25933-5

Samorani, M., & Laganga, L. R. (2015). Outpatient appointment scheduling given individual day-dependent no-show predictions. *European Journal of Operational Research*, *240*(1). https://doi.org/10.1016/j.ejor.2014.06.034

Srinivas, S., & Ravindran, A. R. (2018). Optimizing outpatient appointment system using machine learning algorithms and scheduling rules: A prescriptive analytics framework. Expert Systems with Applications, 102, 245–261. https://doi.org/10.1016/j.eswa.2018.02.022

World Health Organization. (n.d.). *Cancer*. World Health Organization. https://www.who.int/health-topics/cancer#tab=tab_1

Zacharias, C., & Pinedo, M. (2014). Appointment scheduling with no-shows and overbooking. *Production and Operations Management*, *23*(5), 788–801. https://doi.org/10.1111/poms.12065