MIXING SCHEDULED PATIENTS WITH WALK-IN PATIENTS AT COLLABORATIVE EMERGENCY CENTRES

by

Jacob Roland Wing

Submitted in partial fulfillment of the requirements
for the degree of Master of Applied Science

at

Dalhousie University
Halifax, Nova Scotia
December 2017

Table of Contents

List of Tables

List of Figures

Abstract

The Collaborative Emergency Centre (CEC) care model was initiated by the Nova Scotia Department of Health and Wellness in 2010. The CEC model of care appears to be a promising way to reorganize and improve emergency care for rural Nova Scotians. In this thesis, the focus is on improving operations of the daytime CEC where providers see patients who are scheduled, who walk-in, and who return from the nighttime CEC. The models developed in this thesis support the physician in determining how many patients to schedule and when to schedule them. The appointment scheduling system considers the unique characteristics of CECs. The number, and pseudo-optimal placement of scheduled appointments is determined with a simulated annealing procedure that uses computer simulation as the method of performance evaluation.

List of Abbreviations and Symbols

Symbol	Description
p	Patient (scheduled or walk-in)
d	Replication # (day)
D	Number of replications
t	Slot
T	Number of slots in a day
α_d	Volume of scheduled patients that arrive to their appointment on day d
β_d	Volume of walk-in patients that arrive on day d
$A_{p,d}$	Arrival time of patient p
$S_{p,d}$	Start time of patient p
$L_{p,d}$	Appointment length of patient p
$E_{p,d}$	End time of patient p
θ	End of scheduled day for primary provider
${V^s}_{t,d}$	Volume of scheduled patients in system at start of slot t on day d
$E({V^s}_t)$	Expected volume of scheduled patients in system at slot t
${V^w}_{t,d}$	Volume of walk-in patients in system at start of slot t on day d
$E({V^w}_t)$	Expected volume of walk-in patients in system at slot t
W_d	Total waiting time for all patients on day d
${W^b}_d$	Total waiting time for scheduled patients on day d
${W^w}_d$	Total waiting time for walk-in patients on day d
$E(W)$	Average daily total waiting time for all patients
$E(W^s)$	Average daily total waiting time for scheduled patients
$E(W^w)$	Average daily total waiting time for walk-in patients
U_d	Provider utilization on day d
$E(U)$	Average provider utilization
O_d	Total provider overtime on day d
$E(O)$	Average daily overtime

Acknowledgements

I would like to thank my advisor, Dr. Peter VanBerkel for his expertise and guidance throughout the duration of my studies.

I would also like to thank Barbara Cook, Dr. Rob Bush, and Darla MacPherson from the Lillian Fraser Memorial Hospital for providing me with their expertise in the Collaborative Emergency Centre setting.

I also acknowledge the financial support provided by The Nova Scotia Health Research Foundation.

Chapter 1 Introduction

Outpatient appointment scheduling has long been recognized as a difficult process due to the random nature of the healthcare system [1]. Specifically, the variation in patient appointment lengths and arrival times makes efficient appointment scheduling a challenge. As healthcare costs increase, the desire for increased efficiency and the effective use of resources becomes more prevalent. A relatively new challenge in appointment scheduling involves accommodating walk-in patients [2], whom are unscheduled patients that arrive without notice. This research focuses on developing appointment scheduling procedures that account for traditional appointment scheduling challenges, as well as the challenge associated with walk-in patients. The study is motivated by the Collaborative Emergency Centre (CEC) model of care and was conducted at the CEC in Tatamagouche, Nova Scotia.

1.1. Collaborative Emergency Centres (CECs)

In recent years Nova Scotia (NS) has seen major changes in the delivery of care in rural areas. These changes came in response to "The Patient Journey Through Emergency Care in Nova Scotia" report [3] which found, among other things, six-to-seven week waits for daytime primary care appointments, coupled with underutilized emergency services in NS. These concerns were addressed by implementing a new model of care for these rural areas called Collaborative Emergency Centres (CEC), which were initiated by the Department of Health and Wellness in 2010. CECs use a collaborative team of health providers to expand access to primary care and ensure emergency care is available 24 hours a day, seven days a week. The CEC model is used in place of the traditional physician-led emergency department (ED) model, in areas where the volume of patients or the availability of physicians is insufficient. CECs are physically located in emergency department space of existing hospitals and operate in both a nighttime and daytime mode. At night, a nurse and paramedic are on site with access to an on-call physician who can be consulted by telephone. During daytime hours, the CEC is operated by physician and nursing staff with primary care appointments offered from 8am-8pm. The daytime CEC operates like a typical primary care clinic with patients scheduled along with unscheduled arrivals treated as "walk-ins." This overview of how patients flow through CECs can be seen in Figure 1.

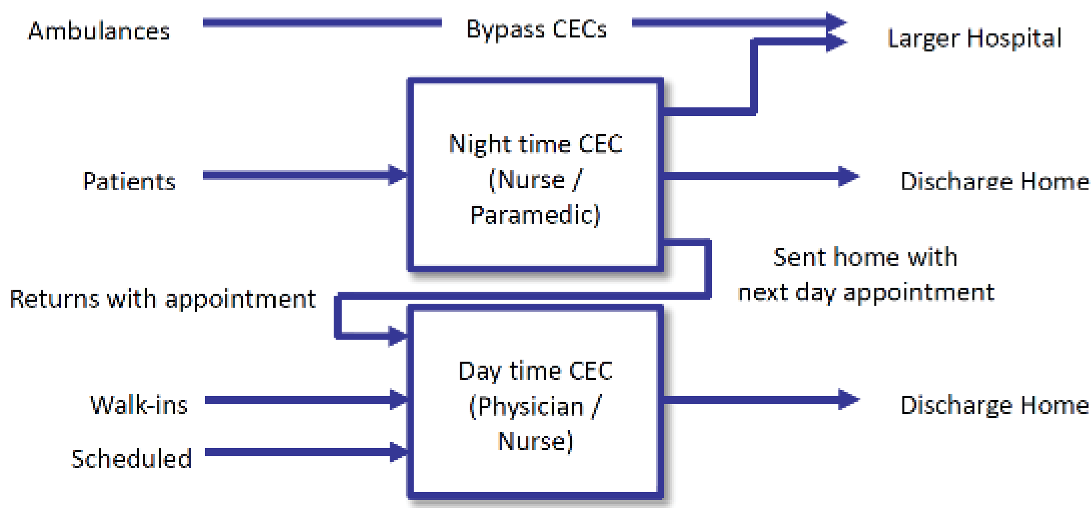

Figure 1: Collaborative Emergency Centre Patient Flow [4]

The objective of a CEC is to provide better access to primary care and improve physician utilization. A recent government sponsored review of the CEC care model has found it to be a success [5]. This review suggests that the CEC program is meeting the goals of increasing primary care access, reducing unplanned closures of overnight emergency care, and reducing the physician cost of providing overnight care. However, a number of challenges related to operational decision making remain. Addressing one of these challenges and optimizing patient scheduling within the daytime CEC is the focus of this research, as will be detailed in Section 1.3

1.2. CEC Scheduling

To achieve the desired improvement in primary care access, the daytime CEC model should guarantee same-day appointments to walk-in patients and returning patients who visited the nighttime CEC. Walk-in patients alone do not constitute a full day's work for the daytime CEC. As such, additional patients (called 'booked patients') are scheduled to fill the clinic. Booked patients include patients who have scheduled an appointment in advance of the previous night (i.e. weeks/days in advance), as well as overnight return patients who received an appointment the previous night. Considering that the average rural hospital in Nova Scotia is seeing 1.3 patients per night [3], the majority of patients booked throughout the day are going to be booked in advance.

Knowing how many booked patients to schedule to ensure same-day access to walk-in patients is a complicated decision. Scheduling too many booked patients means there may be insufficient capacity to deal with walk-ins, while scheduling too few patients means the daytime CEC may be left idle and underutilized. This concern was identified by a focus group participant who stated that the daytime CEC did “not want to book a full day of appointments, because of the unpredictability in the number of outpatients coming through", as well as another participant who stated “same-day / next day appointments [are] not always available” [5].

1.3. Objective

The objective of this study is to design CEC specific patient scheduling methods to ensure that, during daytime operations, CECs reserve sufficient space for walk-in patients, while at the same time appropriately utilize resources. This will be accomplished by developing an appointment scheduling system that indicates a suitable volume of booked patients to schedule in addition to the clinic’s walk-in demand, and when throughout the day to schedule these booked patients. The primary goal of this study is to balance three conflicting objectives: 1) volume of booked patients scheduled, 2) the waiting times of patients, and 3) the overtime scheduling of the care provider(s). This will involve developing a method to evaluate the performance of any given clinic schedule template, as well as using this evaluation tool to determine an optimal clinic scheduling template, i.e. an appointment book that indicates the volume and time throughout the day to schedule booked patients.

1.4. Outline

Chapter 2 will outline the literature on outpatient scheduling and the various techniques used to address the variety of environmental factors unique to each clinic. Chapter 3 will introduce the methodology used throughout this thesis. Chapter 4 describes the context of the case study setting. Chapter 5 demonstrates the application and results of the model presented in Chapter 3. Chapter 6 introduces the tool that was developed for stakeholders to use such that a variety of evaluation methods can be performed independently. Finally, Chapter 7 concludes with a discussion of the completed work, research contributions, and opportunities for future research.

Chapter 2 Literature Review

Outpatient scheduling in healthcare has been of significant interest to researchers and hospital administrators alike since the early work of Bailey [6] and Lindley [7]. The goal of these studies is to develop an appointment system which uses a combination of scheduling rules, control parameters, and mechanisms to determine how patient appointments are scheduled. A comprehensive survey of research on outpatient appointment scheduling can be found in the study conducted by Cayirli and Veral [8], which presents a general problem formulation, model considerations, and the methodologies used in current literature. Gupta and Denton [2] provide a detailed state of the art in the design of appointment systems and also indicate opportunities for industrial engineering and operations research models to be applied to this type of problem.

The general problem being addressed in outpatient appointment scheduling is to assign patients to a specific provider P_p during a well-defined clinic session $j = 1, 2, 3, \dots, J$. Each clinic session is divided into T intervals that are q_t minutes in duration, leading to a clinic duration of $\sum_1^T q_t$ minutes. The length of each interval may vary throughout the session and is called the appointment interval. The number of patients assigned to an identical interval n_t may also vary, and is called the block size.

The following subsections will detail the general rules that are developed when designing an appointment system. Rules that have been studied in the literature will also be detailed.

2.1. Appointment Rules

Appointment rules can be described as the combination of block-sizes, and the length of the appointment intervals. The block-size represents the number of patients that are scheduled to a given block. The appointment interval is the time between two successive appointment start times. Both the block-size and appointment interval can be fixed or variable throughout the clinic session. The following subsections will detail appointment rules that have been studied in the literature using various block-size and appointment interval combinations.

2.1.1. Single-block Rule

This rule schedules all patients to a single block such that they arrive simultaneously at the start of a clinic session. The single-block system puts an emphasis on reducing the provider's idle time by having a consistent backlog of patients ready to be served. This, however, leads to excessive wait times for many of the patients. This type of scheduling system was common prior to early work conducted in outpatient appointment scheduling in the 1950's. Studies conducted by Bailey [6], Lindley [7], and Bailey and Welch [9] introduced the shift towards assigning patients individual appointment times spread throughout the day.

2.1.2. Fixed-block/Fixed-interval

Fixed-block/fixed-interval appointment systems assign the same number of patients to equal length intervals throughout the day. This can include an individual-block system that assigns every patient a unique appointment time, or a multiple-block appointment system, which schedules *n* patients to each block in the session. Soriano [10] developed a Two-at-a-Time appointment system where patients are called two-at-a-time with intervals set equal to double the mean service time.

2.1.3. Variable-block/Fixed-interval

Varying the block sizes throughout the clinic session while keeping the appointment intervals constant allows for added flexibility by strategically adding to and limiting the system load. This can be done in order to better match the patient demand with resource supply by smoothing the workload. A commonly studied system of this type uses the Bailey-Welch Rule. This rule is a variation on the individual-block appointment system, except that two patients are scheduled to the initial block of the clinic session. Bailey and Welch [9] concluded that for one doctor this rule performs the best on a trade-off between patient waiting time and provider idle time. Cayirli and Gunes [11] aim to smooth variability in patient demand by recognizing seasonality in unscheduled arrivals. Varying block sizes between 0, 1, and 2 patients to each slot was done to compliment this seasonality. Further investigation into variable-block/fixed-interval systems have been conducted by Fries and Marathe [12], as well as Liu and Liu [13].

2.1.4. Fixed-block/Variable-interval

This scheduling rule assigns a fixed number of patients to each block, (often individual), and varies the appointment interval length throughout the day. This was studied by Ho and Lau [14] who tested eight different variable-interval scheduling rules under varying environmental conditions. It was found that of the rules tested, the best performance gains were produced by increasing the length of appointment intervals towards the latter part of the session.

2.2. Environmental Factors

Models that have been used in the literature range greatly in complexity. The inclusion of a large number of environmental factors allows the model to represent reality as closely as possible. However, the complexity this adds to the study makes many models intractable. The environmental factors that affect appointment scheduling are detailed in the following subsections.

2.3.1. Number of Providers

In the literature on outpatient scheduling, a traditional assumption is that the clinic is served by one provider, while in reality these facilities often employ multiple providers. If waiting time was the only metric of interest, it would be most efficient to pool these providers using a single common queue for all patients. Schaak and Larson [15] use this common resource pool approach by assigning patients to the first available provider given that a threshold number of providers are left available for higher priority patients. Liu and Liu [13] also use a multi-server system that assumes providers are identical, assigning patients to providers solely based on availability. Many clinic settings will find this pooling of providers undesirable due to the importance of continuity of care that is achieved through a one-to-one provider-patient relationship. This leads to studies often incorporating independent queues for each provider ([16], [17]).

2.3.2. Patient/Provider Punctuality

Studies often assume that patients and providers are available at the time of appointment. In reality, this is not always true and can lead to disruptions in the appointment schedule. Klassen and Yoogalingam [18] consider the punctuality of providers and find that as lateness increases, appointment blocks should be shorter and pushed later into the session. Other studies have fit theoretical distributions to represent the punctuality of patients [19], [20], [21].

2.3.3. Patient No-Shows

Patients not arriving for their given appointment is another source of schedule disruption. This type of patient is often called a no-show. This is a commonly modelled phenomenon and is often represented as a no-show probability that ranges from 5 to 30 percent depending on clinic observations [19], [21]. A common remedy to disruptions in the schedule caused by patient no-shows is to overbook appointment blocks. This also leads to a risk of increasing patient waiting time, causing a trade-off between patient waiting time and provider idle time to be made. Samorani and LaGanga [22] aim to overbook appointments in an optimal way, using no-show predictions that are dependent on the individual appointment characteristics and the appointment day.

2.3.4. Provider Interruptions

Few studies have analyzed the impact provider disruptions have on appointment scheduling [23] . These disruptions may include charting, phone calls, and aiding other staff members. Klassen and Yoogalingam [18], [23] use a simulation optimization approach to determine the best schedule changes as the level of provider interruption changes. It was observed that as interruptions rise, appointments in the middle of the session should be made longer.

2.3.5. Service Time Distribution

The service time distribution is a representation of the time providers consult with patients. Numerous distributions have been used to model service time in the literature. Distributions that have been used in studies include Gamma ([6], [9], [21]), lognormal ([24], [11], [25]), uniform ([13], [14]), Weibull ([13]), and negative exponential ([12]). Empirical data is often collected in order to determine the theoretical distribution that is appropriate for the given clinic setting.

2.4. Modelling Techniques

Methodologies used throughout the literature to model and evaluate a given appointment system are either simulation or analytically based. These two approaches will be detailed in the following subsections.

2.4.1. Simulation

Simulation modelling is a common method used to evaluate appointment system designs and control policies. These models allow for complex queueing systems to be represented while also considering many real-world variables, such as those described in Chapter 2.2. Simulation allows for the complex nature of an outpatient clinic to be captured and represent reality as close as possible. These models are used to evaluate various appointment schedules and scheduling control policies such that informed decisions can be made.

The earliest work in appointment scheduling conducted by Bailey [6] was evaluated using a manual Monte Carlo simulation technique. After testing various sizes for the initial block and appointment interval, it was determined that an initial block of two patients and individual successive appointments led to a good balance between patient waiting time and provider idle time.

A number of studies use simulation to evaluate the placement of open blocks for unscheduled patients. Su and Shih [26] tested four alternative sequencing rules for scheduled and unscheduled appointments, with each being evaluated based on patient throughput and waiting time. Results indicated scheduled and unscheduled appointment slots should be alternated in sequence. Klassen and Rohleder [24] developed a simulation model to test various scheduling rules and appointment blocks to be left open for unscheduled patient arrivals. It was indicated that a trade-off must be made between the various performance measures. Open blocks for unscheduled arrivals left earlier in the session lead to lower patient waiting time and a worse percentage of urgent clients seen, while open blocks left later in the session led to lower provider idle time and more urgent patients being seen. Therefore, the position of the urgent slots throughout the day should be chosen based on the goals of the clinic and the performance metrics that are most highly valued. Rising et al. [16] attempt to smooth demand on physicians by intuitively placing appointments during periods of low unscheduled arrival demand. A Monte Carlo simulation was used to evaluate the appointment system design, which showed a 13.4 percent increase in the number of patients seen by physicians. Cayirli and Gunes [11] use simulation optimization to develop a variable block system that optimally assigns either 0, 1, or 2 patients to a block.

2.4.2. Analytical Studies

Analytical approaches to the study of appointment scheduling include optimization models using queuing theory and mathematical programming methods. The complex nature of outpatient scheduling often makes analytical approaches intractable. Therefore, analytical studies often reduce complexity by making certain assumptions or ignoring certain environmental factors.

Qu et al. [27] use a Markov Decision Process to develop an optimal walk-in patient admission policy. The policy evaluates whether or not to admit a walk-in patient and when that patient should be seen based on the current state of the system. They found that when walk-in patient arrival rates are not greater than 20% of service rate, all walk-in patients should be admitted. If this condition does not hold true, walk-in patients should be admitted when the number of slots remaining is greater than the expected number of patients needed to be seen. Schaak and Larson [15] introduce an N server priority queuing model, where N_i is the threshold level for priority i patients. Priority i patients are served if and only if fewer than N_i providers are occupied. This type of system ensures that high-priority patients receive timely access to providers. Since patients are assigned to any available provider once the threshold conditions are met, this model does not account for patient-provider preferences and continuity of care. Brahimi and Worthington [28] develop a model for multi-server queues with time-dependent arrival rates and discrete service time distribution. The queuing system applied is an inhomogeneous Markov chain in discrete time, using computer programming to calculate time-dependent state probabilities such that idle time and patient waiting time metrics could be obtained. It was determined that the queueing model could be used to show hospital decision makers that improvements to appointment systems are possible, and to advise on what those improvements should be.

2.5. Performance Measures

The goal of an appointment system is to optimize some measure that is a proxy for desired clinical performance. Metrics commonly used in literature include the minimization of patient access time (measured in days waiting for an appointment), patient waiting time (measured in minutes waiting in the clinic to start an appointment), provider idle time, and provider overtime. Often the measure of performance is a combination of these factors, as shown by Peng, Qu, and Shi [3] who aimed to minimize the weighted sum of patient waiting time, provider idle time, and provider overtime.

There are also studies that attempt to maximize some clinical measure of performance. Kortbeek et al. [4] develop an optimal appointment system that maximizes the number of unscheduled patients served on the day of arrival. Su and Shih [26] evaluate various blocks to leave open for unscheduled arrivals based on maximizing patient throughput.

Cost measures can often be conflicting such that a trade-off level must be determined by decision makers based on their priorities for clinic performance. For example, provider idle time could be minimized to zero if a large number of patients were booked into identical blocks. This would however lead to large wait times for patients. Therefore, a trade-off must be made between provider idle time and patient waiting time.

2.6. Managing Unscheduled Arrivals

Appointment scheduling is further complicated in facilities that serve both scheduled appointments and unscheduled arrivals. These unscheduled arrivals are generally called walk-ins. Many clinics do not accommodate walk-in patients and only serve patients with an appointment. As such, many studies in the outpatient scheduling literature do not model walk-in patient arrivals. For facilities that accept walk-ins, issues that must be addressed include randomness of walk-in patient arrivals, different appointment lengths for scheduled and unscheduled appointments, and patients having different needs and priorities. Decision makers must determine how many scheduled patients to book given how many appointments are currently booked, how many are expected to be booked tomorrow, and how many walk-in arrivals are expected. The position of these scheduled appointments must also be determined. The main challenge in scheduling appointments while considering walk-in patients is balancing the trade-off between serving walk-in patients promptly, without increasing the time scheduled patients must wait.

Rising et al. [16] use a two-step approach to smooth the demand for physician services by determining the arrival patterns of walk-in patients and scheduling appointments during periods of low walk-in demand. The first step smoothes the demand by day of the week based on an analysis of historical trends in daily arrival patterns. The second step uses a Monte Carlo simulation model to evaluate the effects of alternative sequencing rules used to smooth the demand across hours of the day. Intuitively choosing appointment periods that compliment known hourly arrivals of walk-in patients led to an increase in the number of patients seen by the physician, as well as a decrease in patient waiting time. Kim and Giachetti [29] develop a stochastic mathematical overbooking model (SMOM) which considers the probability distribution of walk-in patients. This aids in obtaining the optimal number of appointments to be scheduled, such that expected total profits are maximized. Qu et al. [27] develop a finite-horizon MDP model to optimize the walk-in patient admission policy that determines whether or not to admit a walk-in patient and when walk-in patients should be seen.

2.6.1. Arrival Patterns

The complexity of outpatient scheduling is further complicated by the random nature of walk-in patient arrivals. This disruption can be minimized if appointment systems recognize that seasonality in walk-in patient arrivals exists. This includes demand patterns for month of the year, day of the week, and hour of the day. Patient demand seasonality is a clinical characteristic that has not been studied extensively in the healthcare operational research literature [2]. Au-Yeung et al. [30] found walk-in arrivals exhibited a strong 7-day (weekly) seasonality. Rising et al. [16] conducted one of the earliest studies that attempted to take advantage of this information and designed an appointment system that compliments this seasonality of patient arrivals. It was found that scheduling appointments to complement the day of week and time of day arrival patterns helps smooth the patient flow and improves clinic efficiency.

2.7. Discussion

A large amount of work has been conducted in the field of outpatient appointment scheduling, however, there are still challenges and opportunities for future work. Cayirli and Veral [8] recognize that the majority of studies focus on a specific clinic setting, which leads to a lack of generalizability. Developing models that can be tailored to unique clinic settings would be of value such that environmental factors are appropriately represented for the specific clinic being addressed. Many studies do not include all environmental factors in order to simplify their models such that they are tractable. Along with this, there is also a need for models that provide a realistic representation of an outpatient clinic. Models will be far more representative of realistic operating conditions with the inclusion of many environmental factors that are based on data from specific clinics. Empirical data can be used to determine realistic distributions and rates for factors such as walk-in patient arrivals, provider and patient unpunctuality, and no-shows.

A lack of studies surrounding a realistic approach to multiple providers has also been seen. Literature often assumes either a single provider, independent queues for multiple providers, or uses multiple servers as a common pool of resources that can serve any patient. The literature is lacking a model that considers a mix of these types of systems. It is important to adhere to patient preferences and allow for continuity of care by having a one-to-one patient-provider relationship, however, clinics may also have a set of doctors that are available to serve any arriving patient during specific sessions. This mixed type of clinic setting that allows for both specific patient-provider queues as well as a single queue that can be served by multiple providers has not been studied. This type of model adds flexibility while also allowing continuity of care to be respected.

Chapter 3 Methodology

To improve the daily operations of the Collaborative Emergency Centre (CEC), the best schedule template to compliment the demands and environmental factors of the clinic was determined. First, a means to evaluate the performance of any given schedule template was developed using a simulation. Next, a model was developed that determines the "best" schedule template for the given clinic setting. The models used to evaluate schedule templates, and determine a pseudo-optimal schedule template are described in the following sections.

3.1. Evaluation Model

Throughout the operation of a clinic day at a CEC, patients tend to spend some amount of time in a queue waiting to see a care provider. Queues are caused by an imbalance between supply and demand, and/or randomness in patient arrivals and patient throughput. In literature, queueing theory is often used to model the nature of queues. However, due to complexity, high variation, and the possibility of an imbalance between supply and demand, queueing theory is not ideal in many complex health care settings [8].

In place of queueing theory researchers often turn to simulation. An advantage of simulation modeling over queueing theory is the ability to model complex queuing systems with greater accuracy and allow for the inclusion of a wide variety of environmental factors. This was shown in the work conducted by Kolker [31] that uses quantitative examples to demonstrate the practical limitations of queuing theory approaches and the powerful, flexible, and informative nature of simulation modelling as a methodology. For these reasons simulation modelling was chosen for this study as a means to evaluate the performance of a specified schedule template.

3.1.1. Design Approach

Simulation modelling allows for the testing of large scale changes to clinic operations, without needing to physically alter and disrupt any part of the actual clinic. Simulation modelling is a mathematical approach which allows for the longterm behaviour of the clinic to be observed, from which performance measures of interest can be computed. Due to the stochastic nature of simulation models, these performance measures are estimates of system performance.

The simulation model was developed to evaluate and compare the performance of alternative schedule templates while accounting for various clinic specific factors, such as walk-in patient arrival patterns, patient priorities, multiple providers, and variable appointment lengths. A schedule template is defined as the volume of patients booked in the day, and the times throughout the day at which these booked patients are to be scheduled.

The model allows for analysis of variable-block/variable-interval appointment rule systems. This means that any number of booked patients can be assigned to a given time slot, and the interval between booked appointments can be any length of time (later this is limited to 5 minute intervals for computational reasons). The schedule template is intended to be followed for a given clinic day in perpetuity. An example schedule template that books 14 patients can be seen in Figure 2. Variable-block/variable-interval appointment rules have not been studied extensively in literature.

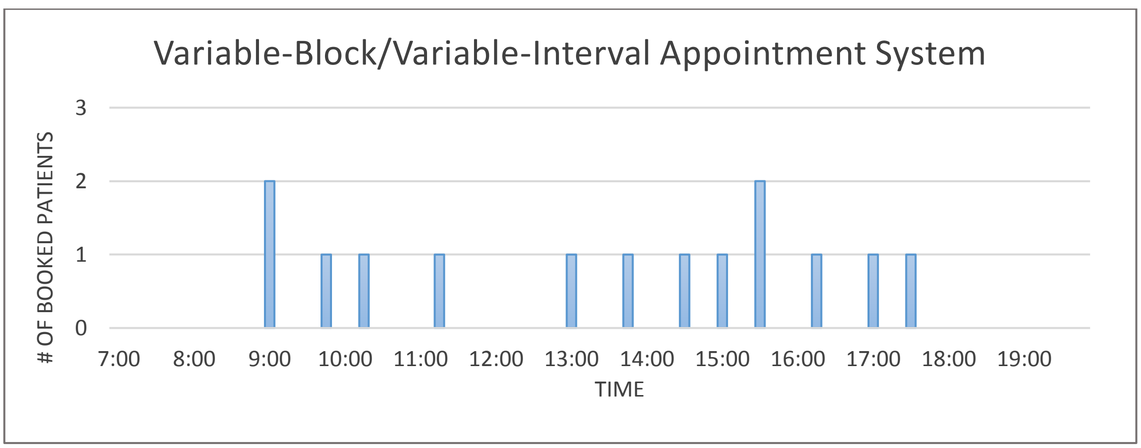

Figure 2: Variable-Block/Variable-Interval Appointment System

To develop the simulation logic, a conceptual model of the clinic was constructed through discussion with clinic staff including the site manager, care providers, and administrative staff. These discussions were further supported by observing the clinic and reviewing datasets. Using this conceptual model and data drawn from Nova Scotia Health Authority databases, a computer simulation was developed in Python 3.6. The model was then tested and validated using a number of techniques described in [32], as reviewed in Chapter 3.4.

3.1.2. Conceptual Model

The clinic day is divided into T equal length slots, with each time slot indexed by t. The individual slot lengths in this model are 5-minute intervals, such that over a 7am-8pm clinic day, there are $T = 156$ distinct time slots, i.e. $t \in \{1,2, \dots, T\}$. The clinic day operates from 8am to 8pm, however, walk-in patients may arrive to wait any time after 7am. One replication of the simulation model represents one clinic day, where each day is independent of other days in the model. The simulation is run for D clinic days, with each instance of a day indexed by $d \in \{1,2, \dots, D\}$.

The entities modelled in the simulation are patients denoted by p. The patient types modelled include booked patients and walk-in patients. Let α_d represent the volume of booked patients on day d, and β_d represent the volume of walk-in patients on day d. On each day, booked patients are indexed from $p = 1, 2, \dots, \alpha_d$, and walk-in patients indexed from $p = \alpha_d + 1, \alpha_d + 2, \dots, \alpha_d + \beta_d$. Within their own patient types, booked patients and walk-in patients are indexed in ascending order based on arrival time to the clinic. In other words, booked patient $p = 1$ arrived before all booked patients $p > 1$ and walk-in patient $p = \alpha_d + 1$ arrived before all walk-in patients $p > \alpha_d + 1$.

Booked patients are assumed to arrive punctually according to their assigned appointment time, while walk-in patients arrive randomly throughout the day following a known arrival process. The arrival time of patient p on day d is denoted by $A_{p,d}$. Every patient is assigned an appointment length based on an empirical distribution developed from collected clinic data. Let $L_{p,d}$ represent the appointment length for patient p on day d.

Booked patients take priority over walk-in patients, such that when a booked patient is in the system they will always be seen before a walk-in patient. A booked patient, however, cannot pre-empt a walk-in patient that has already begun their appointment. Walk-in patients are assumed to remain in the system until they are seen regardless of the wait.

The resources in the simulation include a primary and secondary provider that are available throughout the clinic day to serve patients. The primary provider serves both walk-in patients and booked patients. Booked patients come from the primary provider's patient population and have been given an appointment with that provider, therefore, only the primary provider may serve booked patients. The primary provider is generally available to serve patients for the majority of the day. The secondary provider, however, is available for limited times throughout the day to assist the primary provider with serving walk-in patient demand. Since walk-in patients do not have an expectation of which provider they will be served by, these patients join a shared queue that is served by both the primary and secondary provider. In the situation where both providers are available when a walk-in patient arrives to the system, the secondary provider will be the preferred provider in order to increase the likelihood that the primary provider will be available if a booked patient arrives.

3.1.3. Model Inputs

The inputs detailed in this section must be provided to the simulation model. These inputs provide information pertaining to the underlying nature of the clinic setting, such that meaningful clinic specific results can be obtained.

Booked Patient Schedule Template

A schedule template is specified by identifying the volume of booked patients scheduled in the day, as well as the times throughout the day they will be scheduled. The simulation model will generate a booked patient at each instance identified by the schedule template, making the assumption that booked patients arrive punctually. The system performance under this input schedule template will be evaluated. An example schedule template can be seen in Figure 3.

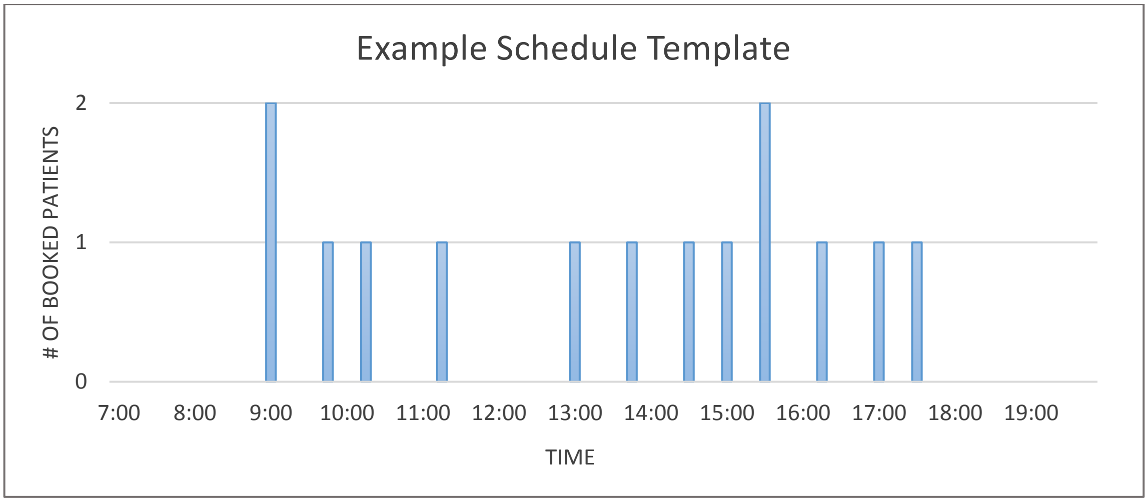

Figure 3: Example Schedule Template

Walk-In Arrival Rate Distribution

In contrast to booked patients that arrive at the times specified by the schedule template, walk-in patients arrive randomly throughout the day. The model requires a time-dependent walk-in arrival rate distribution that is used to randomly generate walk-ins throughout the day. Collected clinic data is used to compute the mean arrival rate for walk-in patients, for each 5-minute interval throughout the day. Like others who model walk-in patient behaviour [33], it is assumed that walk-in arrivals follow a Poisson Arrival process. Using the mean that has been calculated for each 5-minute interval as the parameter for the Poisson Arrival process, a random variate for the number of arrivals during that 5-minute interval is generated. This walk-in arrival rate distribution is explained further in Chapter 4.2.1.

Patient Appointment Length Distribution

The previously described inputs provide the model information on when patients are arriving at the clinic. It is also necessary to know how long they spend with a provider once service begins. The appointment length data collected for this study is aggregated into an empirical distribution. To assign appointment lengths to the simulated patients, a random variate from this distribution for each patient is generated. This patient appointment length distribution is explained further in Chapter 4.2.2.

Appointment Buffer

The time a provider requires between subsequent appointments is called the appointment buffer. This accounts for any notes, charting, etc. that might need to be completed after the patient has left, but prior to the next appointment. This is a constant time measure that is added onto every patient's appointment length. Once this time is added, it is assumed that a provider is able to see the next patient immediately upon completing service of the previous patient.

Provider Availability

The times that the primary and secondary providers are available to serve patients throughout the day must be provided. The primary provider is scheduled to work an 8am-8pm clinic day. However, periods of unavailability may be specified that account for activities such as meals and breaks. During a period of unavailability for a provider, patients may not begin service with that provider. The secondary provider is only available to assist with walk-in demand during set periods throughout the day, with the secondary provider generally unavailable for the majority of the day. Example availabilities for the providers can be seen in Figure 4.

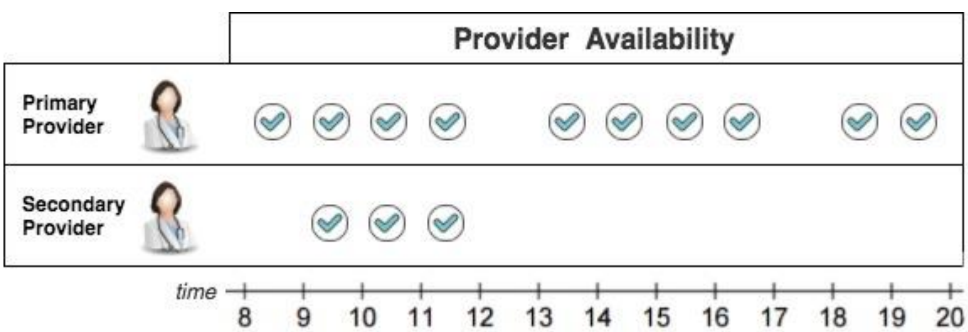

Figure 4: Example Provider Availabilities

3.1.4. Model Logic

The simulation model begins by generating all patients to be seen on a given clinic day, which corresponds to one replication of the simulation. Booked patients are generated according to the schedule template and walk-in patients are generated by producing a random variate of the Poisson distribution associated with every 5-minute interval throughout the day. Appointment lengths are generated for each patient by producing a random variate from the empirical distribution for appointment lengths.

To calculate the system metrics of interest, the appointment start time and appointment end time for each patient must be determined. The start and end time for patient p on day d are denoted by $S_{p,d}$ and $E_{p,d}$ respectively. Ideally patients would begin service at their arrival time, however, provider availabilities, patient priorities, stochastic appointment lengths and walk-in arrivals makes this uncommon.

To demonstrate how the generated patients are processed, the activities and the patient selection logic for the primary and secondary provider are outlined below.

The primary provider can be engaged in one of the following states: taking a break, serving a walk-in patient, serving a booked patient, or waiting for a break/patient arrival. At the end of each of these activities, the provider must decide what their next activity will be. In doing so the provider is logically selecting patients to be seen throughout the day. The daily patient selection logic for the primary provider can be seen in Figure 5.

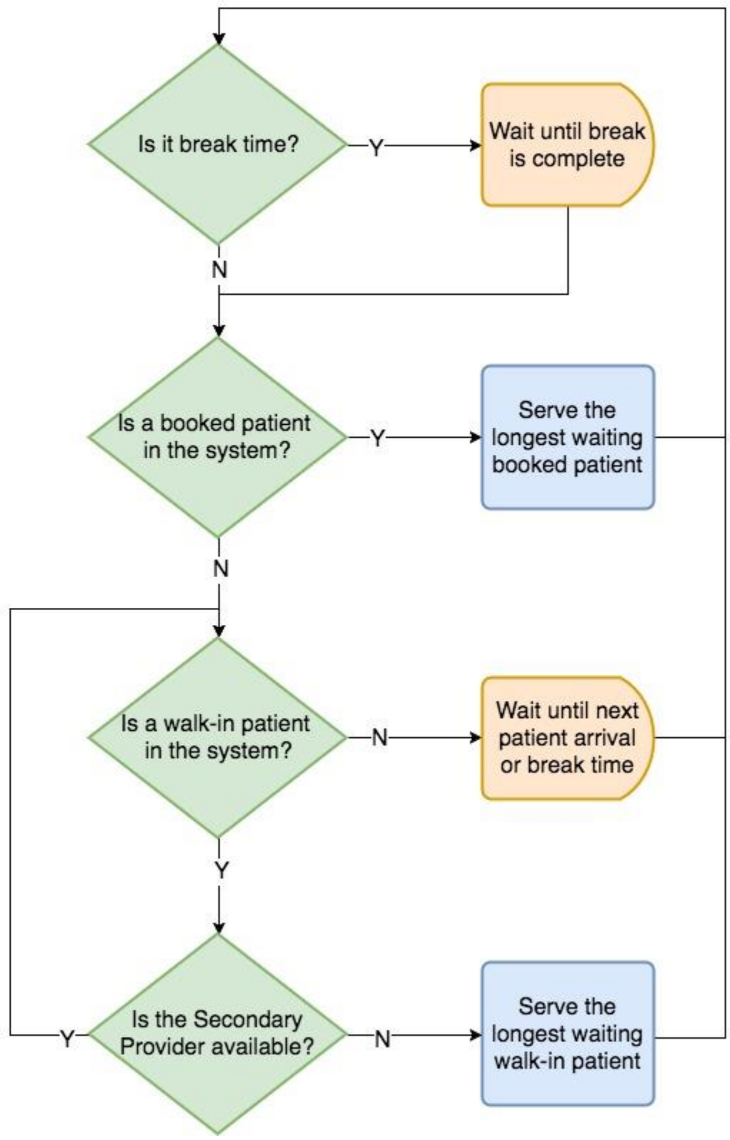

Figure 5: Primary Provider Patient Selection Logic

Once an activity is completed, the primary provider first determines if it is time to take a break, as indicated by the provider availability input. If so, the primary provider takes a break of the specified length. If it is not break time, the primary provider determines whether there is a booked patient in the system. If so, the primary provider will serve the booked patient that has been waiting the longest. The primary provider first checks for a booked patient because this patient type has priority over the walk-in patient type. If there were no booked patients in the system, the primary provider would then check if there was a walk-in patient in the system. If so, they would then determine if the secondary provider is available to serve this walk-in patient, since the secondary provider is the preferred provider for this patient type. If the secondary provider is available, the primary provider will check if another walk-in patient is in the system. If there is a walk-in patient in the system and the secondary provider is not available, the primary provider will serve the walk-in patient that has been waiting the longest. If there are no patients in the system, the primary provider will then wait until the next patient arrival, or their next break time.

The logical walk-through for the secondary provider begins at the start of the day, or once an activity has been completed. The states of the secondary provider include: taking a break, serving a walk-in patient, waiting for a break, or waiting for a walk-in patient arrival. The daily patient selection logic for the secondary provider is shown in Figure 6.

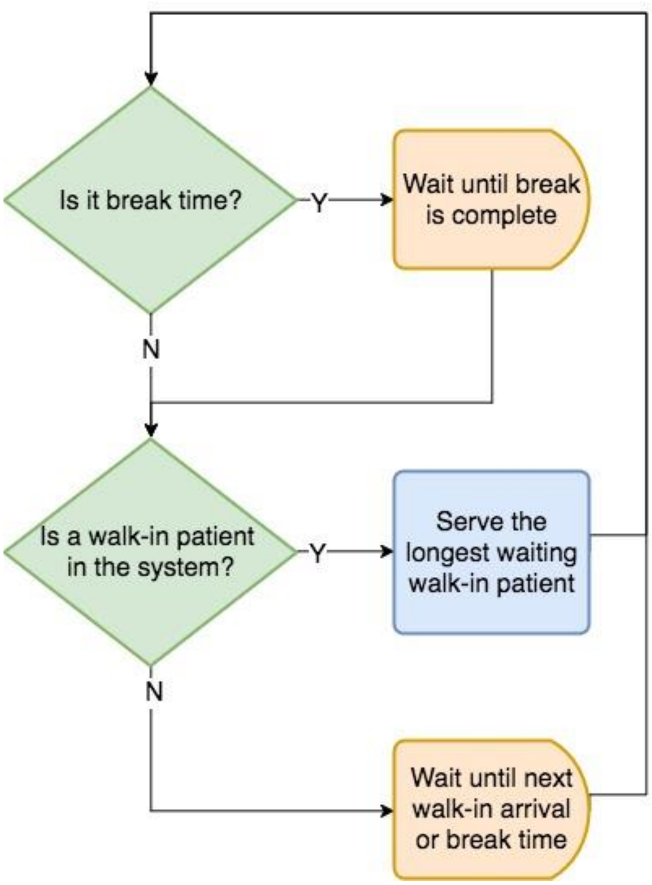

Figure 6: Secondary Provider Patient Selection Logic

Once an activity is completed, the secondary provider first determines if it is time to take a break, as indicated by the provider availability input. If so, the secondary provider takes a break of the specified length. If it is not break time, the secondary provider checks to see if there is a walk-in patient in the system. If so, the secondary provider will serve the walk-in patient that has been waiting the longest. If there is no walk-in patient in the system, the secondary provider will then wait until the next walk-in arrival, or their next break time.

The patient selection logic for the providers in this chapter describes how patients are selected and processed in the system. This in turn allows the model to determine each patient's provider, appointment start time, and appointment end time for service. These start and end times along with the arrival times for each patient is all that is required to calculate the metrics of interest. The simulation parameters and metric calculations will be detailed in the following section.

3.1.5. Model Evaluation

Simulation Parameters

Time Step

The model divides the day into time slots of equal length. For the purposes of this research the time step has been set to 5-minute increments. Using this time step over a 7am-8pm clinic day, as in the model, leads to 156 distinct time slots. The time step value can be increased or decreased. If it is decreased the computation time will increase but will accommodate more refined lengths of time (i.e. time intervals that are not multiples of 5). In this application, the 5-minute time step allows the model to both run in a reasonable amount of time and provide meaningful results.

Run Length

The daytime operation of an outpatient clinic is a finite horizon system. Patients enter the system at the start of the day, and continue to be served until the end of the day, at which point the system empties. Therefore, the natural run length of a replication is one clinic day.

Number of Replications

The reliability of the model results will increase as the number of replications increase. However, running the model too long can be a time-consuming task. The number of replications required to achieve a 95% confidence interval with a 10% relative error for the metric of interest is determined. This will be accomplished using a sequential approach as described by Law and Kelton [32], and demonstrated below.

Let X_i be the measure of a metric of interest, based on the i^{th} replication for $i = 1, 2 \ldots n$. The average of this metric across n replications can be computed by $\overline{X}(n) = \frac{1}{n}\sum_{i=1}^{n} X_i$. An approximate $100(1-\alpha)$ percent confidence interval for the mean is given by:

$$Confidence\ Interval\ of\ Mean: \qquad \bar{X}(n) \pm t_{n-1,\frac{\alpha}{2}}\sqrt{\frac{S^2(n)}{n}}$$

Where $\qquad n = \#\ of\ replications$

$$t_{n-1,\frac{\alpha}{2}} = value\ from\ Student's\ t - distribution$$

$$S^2(n) = sample\ variance$$

$$h(n) = t_{n-1,\frac{\alpha}{2}}\sqrt{\frac{S^2(n)}{n}} = half\ width\ of\ confidence\ interval$$

$$\gamma' = \frac{\gamma}{(1+\gamma)}, where\ \gamma\ is\ the\ preferred\ relative\ error$$

To begin the sequential procedure, the model is initially run with an arbitrary number of replications n_0, which n is set equal to. This initial run of the model allows the point estimate of the mean $\bar{X}(n)$ and half width of the confidence interval $h(n)$ to be computed. If $\frac{h(n)}{\bar{X}(n)} \leq \gamma'$ is true, then the requirement of obtaining a 95% confidence interval with 10% relative error has been satisfied. If not, the value of n is increased and this testing is repeated. This process of increasing the value of n and testing the condition is continued until the condition is satisfied.

Testing for this study, using a relative error of 10 % and a two-tailed confidence interval of 95 %, leads to a required number of replications equal to 160 to achieve this level of precision.

Initial Conditions

The clinic begins empty at the start of every day (replication), and empties at the end of the day. There is no carry over of patients from one day to another, therefore, days are independent. Note that since the clinic system starts from empty each day, there is no need for a warm-up period that is often included to allow the system to reach steady state.

Performance Evaluation Metrics

The performance of a schedule template can be evaluated based on a number of factors derived from the model. These measures include patient waiting time, clinic overtime, and provider utilization. A well performing schedule template provides a high provider utilization while still ensuring sufficiently nominal values for patient waiting time and clinic overtime. The expected state of the system at a given time in the day can also be observed by computing average patient volumes. The following subsections will detail how these metrics can be calculated using the information gathered from patient sequencing. Recall that a day in the model represents one replication.

Patient Waiting Time

Patient waiting time for patient p on day d is denoted by $W_{p,d}$. This measure is computed as the difference between the patient's start time and arrival time.

$$W_{p,d} = S_{p,d} - A_{p,d}$$

The total patient waiting time for day d is,

$$W_d = \sum_{p=1}^{\alpha_d+\beta_d} S_{p,d} - A_{p,d}$$

This could also be evaluated separately for booked and walk-in patients as:

$$W^b{}_d = \sum_{p=1}^{\alpha_d} S_{p,d} - A_{p,d}$$

$$W^w{}_d = \sum_{p=\alpha_d+1}^{\alpha_d+\beta_d} S_{p,d} - A_{p,d}$$

Finally, the estimate of the expected waiting time for all patients, booked patients, and walk-in patients respectively can be computed. These averages across all simulation replications (D) can be computed as shown below.

$$E(W) = \frac{\sum_{d=1}^{D} W_d}{\sum_{d=1}^{D}(\alpha_d + \beta_d)} \qquad E(W^b) = \frac{\sum_{d=1}^{D} W^b{}_d}{\sum_{d=1}^{D} \alpha_d} \qquad E(W^w) = \frac{\sum_{d=1}^{D} W^w{}_d}{\sum_{d=1}^{D} \beta_d}$$

Clinic Overtime

Let O_d denote the overtime for day d. Overtime is experienced when the end time of the final patient of the day occurs after the last time that the primary provider is scheduled to be serving patients. Let $\emptyset$ represent the time that the primary provider is scheduled to finish serving patients. The primary provider is assumed to continue serving patients until all have been served, including time past this $\emptyset$ value if required. Therefore, determining the amount of overtime required for day d can be determined by subtracting the end time of the last patient by $\emptyset$. If the primary provider is scheduled to see patients until $\emptyset = 6$pm, and the end time of the final patient of the day is 7pm, there is a clinic overtime of 1-hour. The final patient of the day will either be the last booked patient or the last walk-in patient. This means that the end of the day will be the maximum of $E_{\alpha_d,d}$ and $E_{\alpha_d+\beta_d,d}$, where $E_{\alpha_d,d}$ and $E_{\alpha_d+\beta_d,d}$are the end times of the last booked patient and walk-in patient respectively. Therefore, daily total overtime can be computed as:

$$O_d = \max\left[0, \max\left(E_{\alpha_d,d}, E_{\alpha_d+\beta_d,d}\right) - \emptyset\right]$$

The overtime required for any given day is now known, therefore, the expected daily overtime can be computed as:

$$\mathrm{E}(O) = \frac{1}{D}\sum_{d=1}^{D} O_d$$

Provider Utilization

Let U_d denote the utilization for a provider on day d. The utilization of the provider refers to the percentage of the day that the provider is serving patients. This includes both time spent directly serving patients, as well as the appointment buffer that is added to account for charting between successive appointments. This time serving patients throughout the day is compared to the total time the provider is available, such that utilization can be computed as:

$$U_d = \frac{total\ time\ serving\ patients\ on\ day\ d}{total\ time\ available\ to\ serve\ patients\ on\ day\ d}$$

The total time that the provider is available throughout the day is predetermined by the inputs for provider availability. The total time that is spent serving patients throughout the day must then be determined. This can be calculated by analyzing each time slot to determine whether there is a patient being served by the provider of interest. Once the provider utilization on each replicated day has been calculated, the expected daily utilization across all replications can be computed as:

$$E(U) = \frac{1}{D}\sum_{d=1}^{D} U_d$$

Slot-Dependent Booked Patient Volume

Let $V^b_{t,d}$ represent the volume of booked patients in the system during slot t on day d. For each arriving booked patient on day d, it can be determined whether they are in the system at slot t by evaluating whether slot t is positioned between the patient's arrival time and exit time. This logic is used to count the volume of booked patients in each slot throughout the day for each day replicated in the model. The expected volume of booked patients in a given slot can then be calculated by taking the average of all replicated days. Let $E(V^b_t)$ represent the average volume of booked patients in slot t, which can be computed as:

$$\mathrm{E}\left(\mathrm{V^b}_{\mathrm{t}}\right) = \frac{1}{D}\sum_{d=1}^{D} \mathrm{V^b}_{\mathrm{t,d}}$$

Slot Dependent Walk-In Patient Volume

Similarly, the volume of walk-in patients in the system at a given time slot t can be determined. Let $\mathrm{V^w}_{\mathrm{t,d}}$ represent the volume of walk-in patients in the system during slot t on day d. For each walk-in patient on day d it can be determined whether they are in the system at slot t by evaluating whether slot t is positioned between the patient's arrival time and exit time. This allows the volume of walk-in patients in each slot throughout the day for each day replicated in the model to be counted. The expected volume of walk-in patients in a given slot can be determined by taking the average of all replicated days. Let $E(V^{\mathrm{w}}{}_t)$ represent the average volume of walk-in patients in slot t, which can be computed as:

$$E(V^{\mathrm{w}}{}_t) = \frac{1}{D}\sum_{d=1}^{D} V^{\mathrm{w}}{}_{t,d}$$

3.2. Optimization Model

In addition to evaluating a single schedule template, the goal of this research is to determine an optimal schedule template that specifies when to schedule booked patients throughout the day as to compliment the arrival of walk-in patients. The simulation model described in Chapter 3.1 allows any given schedule template to be evaluated. To find an optimal solution, complete enumeration could be used to evaluate all possible schedule template combinations. This would quickly become very computationally expensive, even when there are very few booked patients. Therefore, a heuristic procedure is used to find a near optimal solution, which will be referred to as a pseudo-optimal solution. This heuristic procedure requires the number of booked patients to be scheduled as an input, with the goal of determining the best times throughout the day to schedule these patients. If the number of booked patients to schedule in a given day is unknown, this heuristic procedure can be run for various booked patient loads such that the impact on the metrics of interest can be observed. This will provide a means to determine a suitable volume of booked patients to schedule, as well as when to schedule them throughout the day.

Simulated annealing was chosen as a good heuristic procedure to find a pseudo-optimal solution, while also solving much faster than a complete enumeration approach. Simulated annealing was inspired from annealing in metallurgy, which is a technique involving heating and controlled cooling of a material to increase the size of its crystals and reduce their defects [34]. The advantage of using simulated annealing as a global optimization technique is the ability for the algorithm to avoid getting stuck at a local minimum value, as will be described in the algorithm below. This heuristic was also chosen due to its relatively easy implementation, and the high likelihood of providing a pseudo-optimal solution. Since there are further clinic considerations not captured in the model (such as provider preferences), a near optimal starting point is all that is required. Decision makers can tweak this pseudo-optimal schedule to accommodate any additional issues presented. Therefore, optimality for the output schedule template is not required and a pseudo-optimal solution is reasonable. The following subsections detail the simulated annealing algorithm as implemented.

3.2.1. Simulated Annealing Procedure

The simulated annealing algorithm uses a random search procedure that accepts changes that improve the objective function and some changes that degrade it [35]. This process will help avoid getting stuck in a local minimum and provide a thorough search of the solution space. This algorithm will be referred to as the simulated annealing (SA) procedure, and will be detailed in the remainder of this section.

1. Generate a random initial schedule template

The SA procedure begins by generating a random initial solution. For this study, a schedule template is generated with the input number of booked patients scheduled at random times throughout the day. An example initial schedule template with an input of 10 booked patients can be seen in Figure 7.

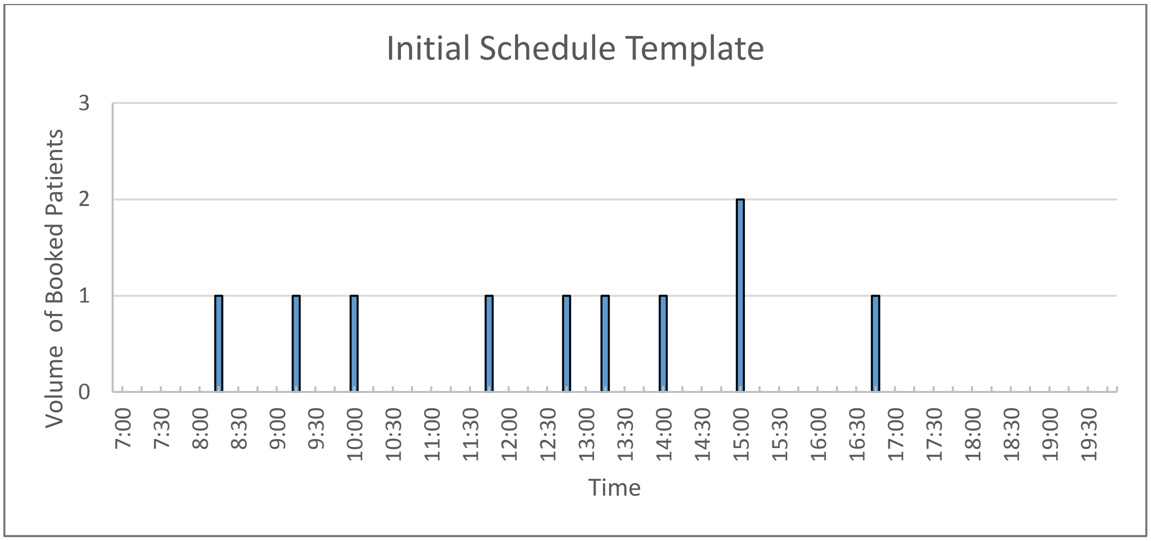

Figure 7: Example Initial Schedule Template

2. Calculate cost of initial schedule using cost function

In order to run the optimization procedure, an objective function must be specified that is to be either minimized or maximized. For example, to evaluate the overall performance of a schedule template, an objective function that is equal to a weighted combination of average patient waiting time and average clinic overtime (both measured in minutes) can be used. Patient waiting time and clinic overtime should remain low, therefore, this cost function is minimized.

$$\boldsymbol{Minimize}: \qquad \boldsymbol{Cost} = \boldsymbol{\alpha} \times (\boldsymbol{PatientWaiting}) + \boldsymbol{\beta} \times (\boldsymbol{Clinic\ Overtime})$$

This step of the simulated annealing procedure evaluates the initial schedule template using the cost function as described. This schedule evaluation is accomplished using the simulation model described in the Chapter 3.1.

3. Generate a random neighboring solution

A new schedule template that is a neighbor to the previous schedule template is now generated. This is accomplished by selecting a booked patient at random and moving them from their currently assigned time slot to a new time slot chosen at random. Continuing on from the example in Figure 7, a random neighboring solution can be generated by moving the patient scheduled at 11:45 to 15:45, as seen in Figure 8.

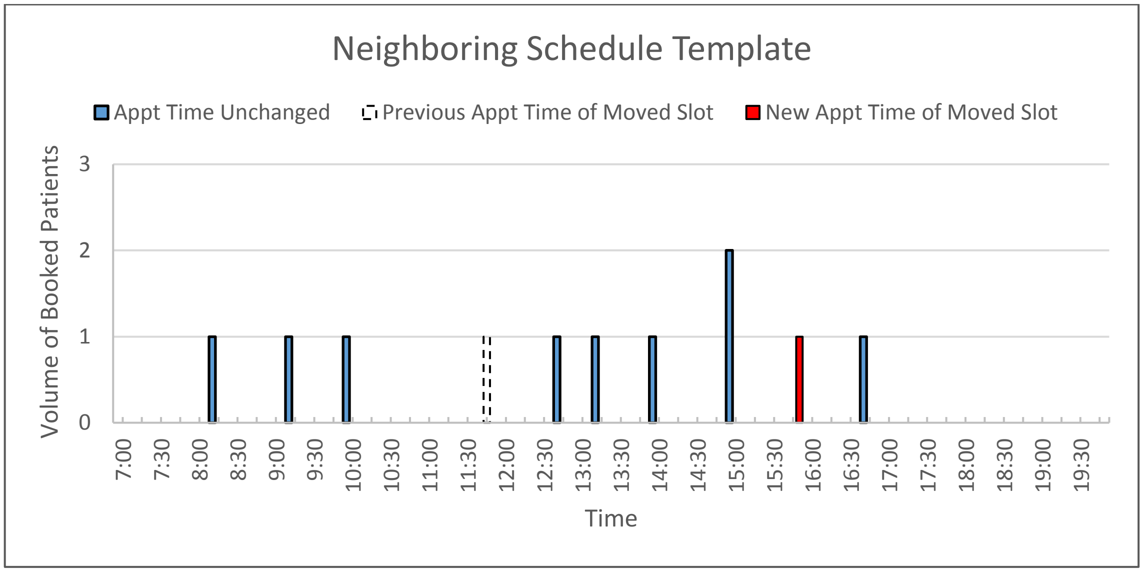

Figure 8: Example Neighboring Schedule Template

4. Calculate the new solution's cost

The new schedule template is evaluated with the simulation, using the same objective function as set in Step 2.

5. Compare cost of old and new schedule template

There are now two similar but distinct schedule templates, and their evaluated cost. It is now necessary to determine whether to pivot away from the old schedule template to the new, or remain with the old. This is accomplished by comparing the cost values. First, if the cost of the new template is less than the cost of the old template, move to the new template since this is indicative of a better performance. If the cost of the new template is greater than the old template, indicating the new template is a worse schedule, move to the new template with some acceptance probability. This is the part of the algorithm that allows the search procedure to avoid getting stuck in a local minimum, and will be further explained in Chapter 3.2.2.

6. Repeat steps 3-5

The procedures described in Steps 3-5 are now repeated by generating random neighboring schedules, comparing to the previously accepted schedule, and accepting/rejecting the new schedule with some probability. This is continued until the selected stopping criteria is met, at which point an acceptable pseudo-optimal schedule template is found.

3.2.2. Acceptance Probability Function

The probability of making the transition from the current state to a candidate state is specified by an *acceptance probability function.* The acceptance probability function takes in the old cost, new cost, and current temperature, as shown in the following equation. If the new cost is less than the old cost, the equation will yield a number greater than 1, and a number between 0 and 1 otherwise.

$$a = e^{\left(\frac{c_{old}-c_{new}}{T}\right)}$$

This acceptance probability is compared to a randomly-generated number between 0 and 1, and if the acceptance probability is larger than the random number, the model switches to the candidate state.

3.2.3. Temperature

In simulated annealing, the temperate T is a global time-varying parameter that is systematically lowered over time. The temperature parameter affects the probability of accepting a new trial solution, with a larger value of T leading to a larger probability of acceptance. The search starts with a large initial temperature, represented by T_0, such that the majority of candidate solutions are accepted. Following each iteration, the temperature is decreased by multiplying T by a constant value between 0 and 1. Simulated annealing does better when the neighbor-cost-compare-move process is carried about for many iterations at each temperature value. For any given finite problem, the probability that the simulated annealing algorithm terminates with a global optimal solution approaches 1 as the search procedure is extended. Therefore, the annealing schedule should be determined such that a compromise is made between producing a good solution, and having an acceptable processing time.

3.2.4. Initial Temperature

Kirkpatrick [36] suggested that a suitable initial temperature T_0 is one that results in an average increase acceptance probability a_0 of about 80%. The value of T_0 will depend on the scaling of the function and, therefore, be problem-specific. It can be estimated by conducting an initial search in which all increases are accepted and calculating the average of the objective increase observed, $\delta = \left(c_{new} - c_{old}\right)^{+}$. The initial temperature is then given by $T_0 = \frac{\bar{\delta}}{\ln(a_0)}$. This process is completed prior to each simulated annealing procedure such that a problem specific initial temperature is ensured.

3.2.5. Stopping Criteria

The temperature parameter is systematically lowered throughout the search procedure. Starting with the initial temperature as described in Section 3.2.4, three iterations of the neighbor-cost-compare-move process are performed for each temperature value. The temperature value is then lowered following the three iterations by multiplying T by a constant that is equal to 0.85 for this research. This search procedure continues until the condition $T < 0.0001$ is met.

3.3. Assumptions

In developing the models used for evaluation and optimization, a number of assumptions have been made. These assumptions have been stated throughout the previous sections and are summarized below.

- Booked patients arrive punctually at their appointment time
- There are no booked patient no-shows
- Walk-in arrivals occur based on a time-dependent Poisson process with arrival rates computed from 1-year of historical data
- Booked patients always take priority over walk-in patients

- Patient appointment lengths are determined by a discrete distribution based on historical data
- Demand is sufficient to ensure that all slots reserved for booked patients in the schedule template are used by booked patients

3.4. Verification and Validation

Verification of a model is the process of confirming that it is correctly implemented with respect to the conceptual model [37]. To complete the verification of the model, it must be shown that the right model has been built given the requirements set forth in Chapter 3.1.4. This was conducted by completing a walk-through analysis of the daily output results of simulation trial runs. This output was compared to manually sequenced patients. This allowed for comparison between the output data generated by the simulation model with the output data expected from the real-world system. Multiple tests were run with varying input parameters, and this debugging procedure was conducted to ensure expected model outputs were achieved. This testing and debugging procedure is a time-consuming but very important part of the modelling procedure. Further verification measures were completed during the development of the simulation model computer program. The model was developed in a modular fashion such that specific functions of the model could be tested to ensure that expected results were being achieved.

Validation is the process of demonstrating that the model is an adequate representation of the actual system being modelled. This means that the model reproduces system behavior with sufficient accuracy to satisfy the objectives set forth. The model has been validated both qualitatively and quantitatively. To validate the model qualitatively, the model, assumptions, flow diagrams, and outputs have been shown to system experts. The experts chosen for this analysis and validation included the site manager, care providers, and administrative staff. According to these system experts, the flows represented in the model are a reasonable representation of how the clinic operates on a daily basis. To validate the model quantitatively, model testing was conducted to compare outputs to expected values. Some of this testing will be detailed in the remainder of this section.

The first test examines whether the results of the metric calculations used in the model make logical sense. This has been completed by examining the waiting time law that indicates waiting time should increase exponentially as utilization increases. This relationship between waiting time and utilization is shown in Figure 9.

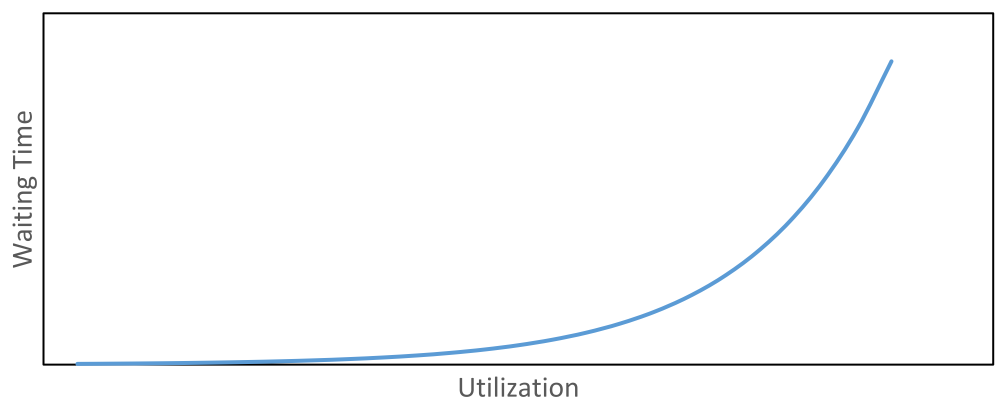

Figure 9: Expected relationship between wait and utilization

The simulated annealing procedure was run multiple times, increasing the number of booked patients after each run. Increasing the number of booked patients is expected to increase provider utilization. Average utilization and average patient waiting time measures were calculated for each of these runs, with the results shown in Figure 10. As expected, the relationship between utilization and waiting time is exponential in nature. This analysis also showed that as the volume of patient demand on the system increases, so too does both waiting time and utilization metrics.

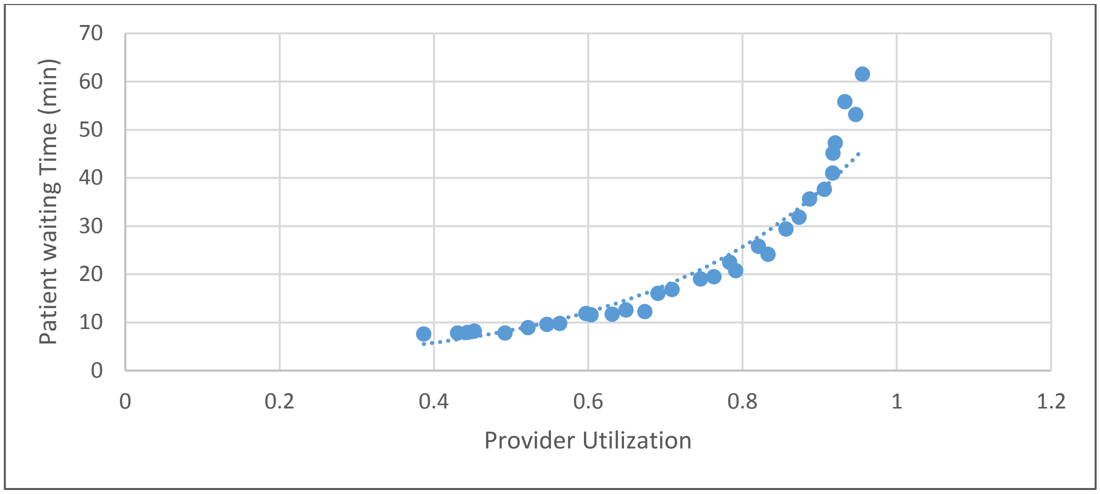

Figure 10: Model relationship between patient wait and utilization

Another quantitative analysis was conducted to ensure that the addition of a secondary provider was leading to expected changes to output values. This was done by running the model with no secondary provider availability, and again with the secondary provider available from 9am-11am. The expectation is that patient waiting time should greatly improve when the secondary provider is available. The model was run for varying loads of booked patients to observe the effects, as depicted in Figure 11. As expected the model outputs for average patient waiting time improve significantly with the addition of secondary provider availability.

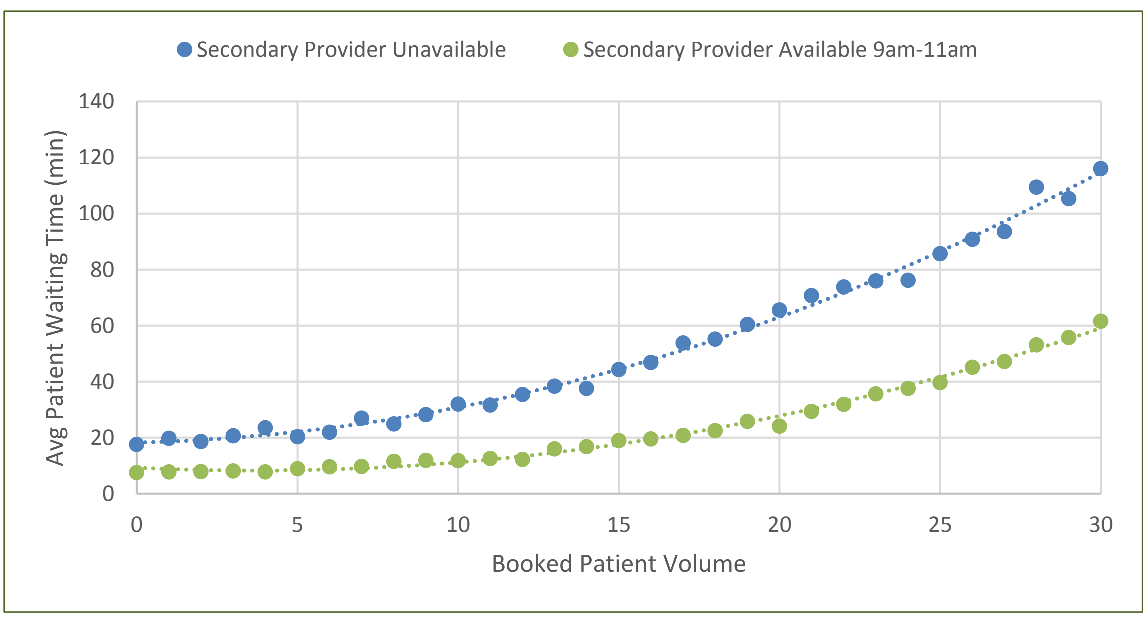

Figure 11: Impact of secondary provider availability

Finally, a run of the simulated annealing procedure was conducted to determine whether the heuristic approach produced improved schedule templates. This was accomplished by comparing the performance of a historically used schedule template to the performance of the pseudo-optimal schedule template produced using the same volume of booked patients.

The historical schedule template seen in Figure 12 was used on a given clinic day in our case study environment to schedule 9 booked patients. Using the simulation model to evaluate the performance of this schedule template leads to an estimate of expected patient waiting time of 28.2 minutes.

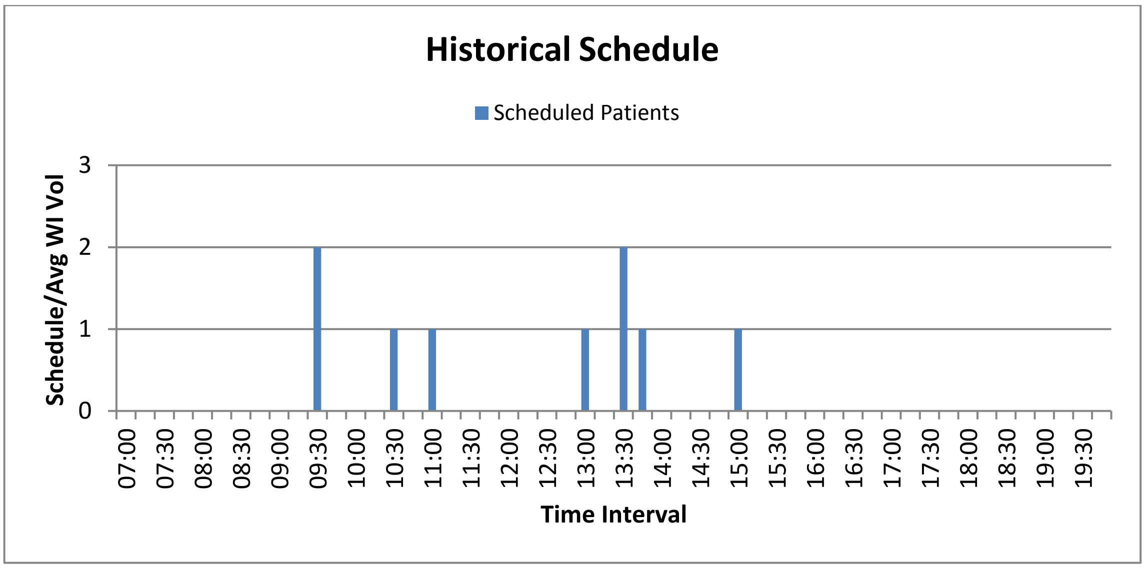

Figure 12: Historically Used Schedule Template

The simulated annealing procedure was then run to determine when throughout the day these 9 booked patients should be scheduled, such that patient waiting time is minimized. The pseudo-optimal schedule template developed from this run of the model can be seen in Figure 13.

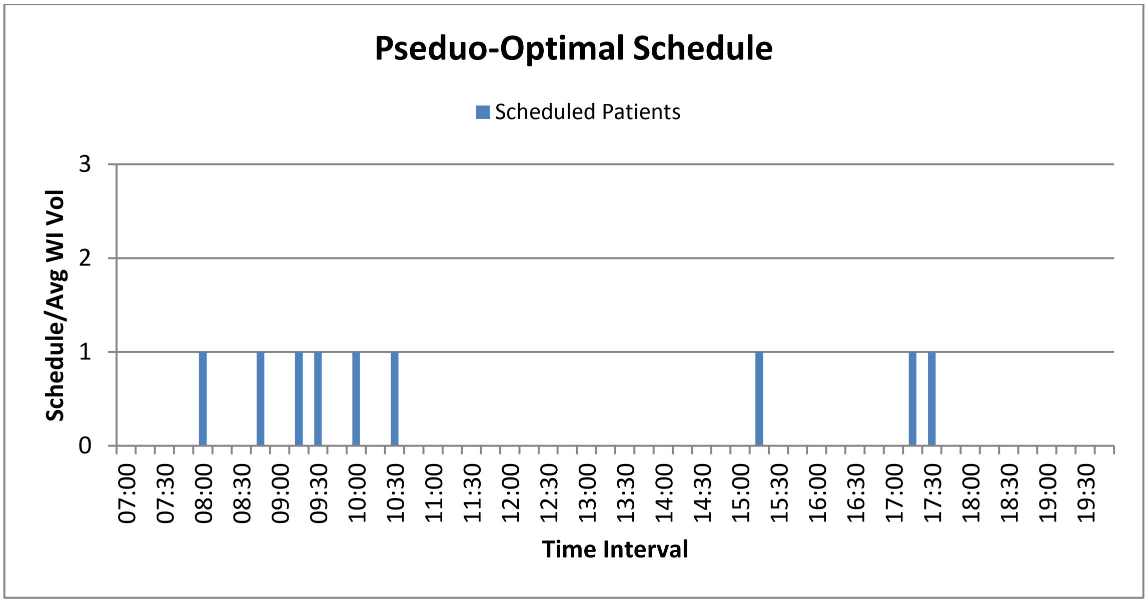

Figure 13: Pseudo-Optimal Schedule

The estimate of expected patient waiting time under the new schedule template is 12.6 minutes. This indicates a 55.3% reduction in average patient waiting time has been achieved by strategically scheduling the 9 booked patients throughout the day.

Chapter 4 Case Study Context

4.1. Clinic Description

The case study for this research has been conducted at the Collaborative Emergency Centre located in the Lillian Fraser Memorial Hospital in Tatamagouche, Nova Scotia. As seen in Figure 14, Tatamagouche is a village located along the south side of Tatamagouche Bay with a population of approximately 750 people [38]. However, the number of people that access primary care through this facility is much higher, as people who live in the surrounding communities also receive care here. This rural hospital facility has operated under the CEC model of care since July of 2012.

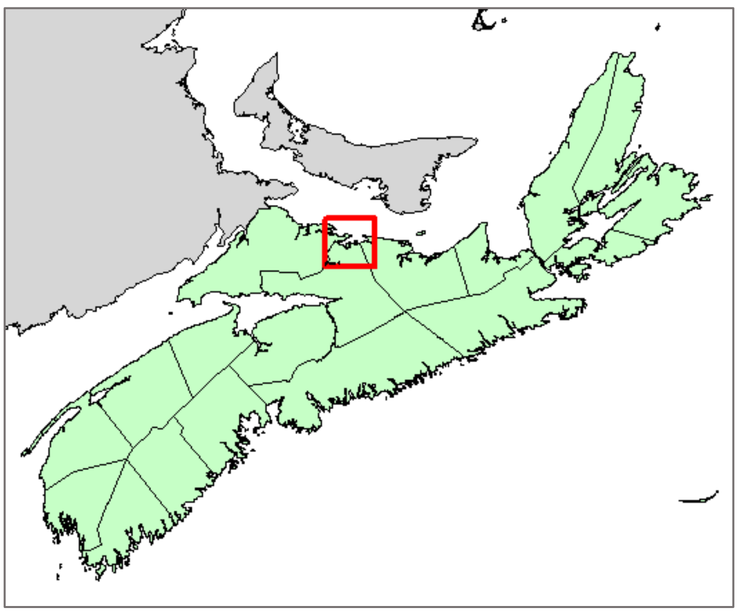

Figure 14: Tatamagouche on Nova Scotia Map

Current appointment methods tend to be variable and completed manually. Booked patients are scheduled on an ad-hoc basis based on the experience and intuition of the scheduling staff. This may be very effective when the scheduler is experienced in this clinic setting, however, this becomes much more difficult when the scheduler is new to the system. This is why a standard schedule template is intended to be an effective method to guide staff when scheduling booked patients throughout the day.

4.2. Clinic Data

Clinic data was collected to provide sufficient inputs to the model. This data allows for clinic specific results to be obtained. The collected data elements and their method of collection are detailed in the following subsections.

4.2.1. Walk-In Patient Arrival Pattern

To generate walk-in patients in the model, information must be provided to indicate when throughout the day these patients are arriving. The data used to determine arrival patterns for walk-in patients were gathered from Nova Scotia Health Authority data sources. The data indicates the registration time for all arriving walk-in patients for a 1-year period from April 2016-March 2017. The time that walk-in patients are registered in the system is used as a proxy for arrival time, since the model assumes that patients are ready to be served directly at their time of arrival. Figure 15 shows the average walk-in arrival rates throughout the day for each weekday that sees a full day of walk-in patients, with results grouped in 15-minute time intervals. It can be seen that there is a consistent trend throughout the day across weekdays. Friday is excluded in this analysis since walk-in patients are not served for the full day.

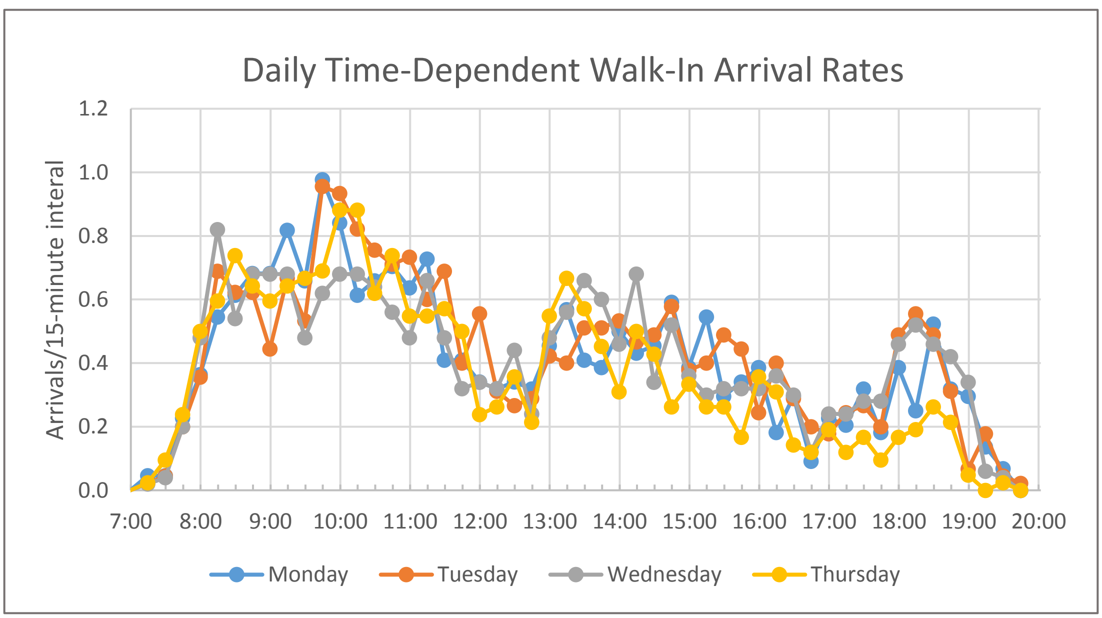

Figure 15: Daily time-dependent walk-in arrival rates

Figure 16 shows the average rates across days Monday-Thursday to clearly illustrate the general trend for when walk-in patients arrive throughout the day. The average daily volume of walk-in patients is 21.82 with a standard deviation of 4.67.

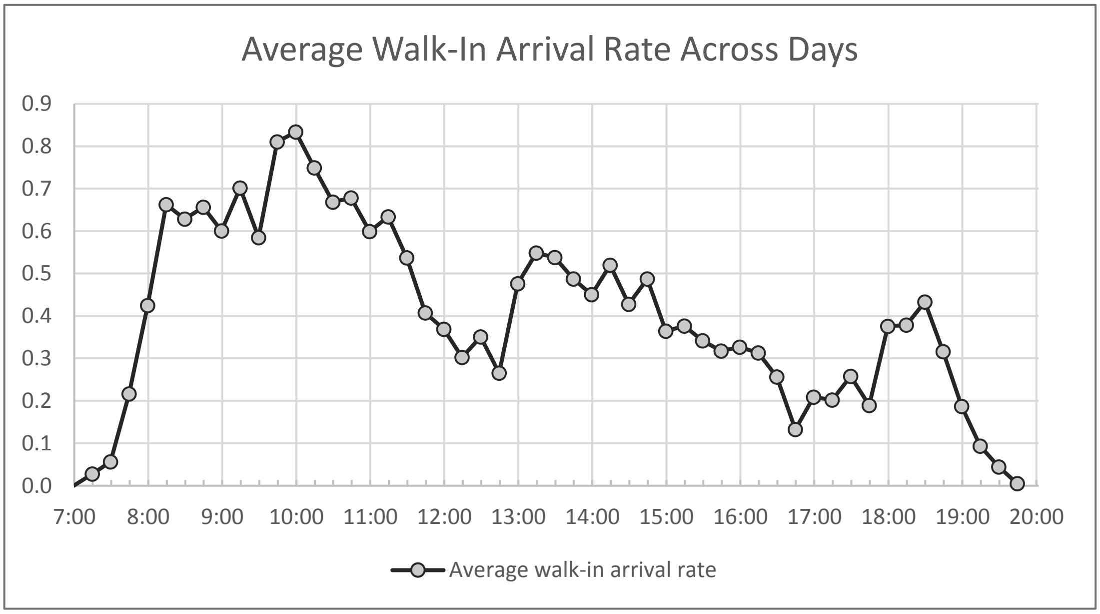

Figure 16: Monday-Thursday average walk-in arrival rate

4.2.2. Appointment Length

Appointment length data was manually collected for all patients over a 2-week period in February 2017. This was accomplished by developing a patient tracker tool. The paper based tool was given to each patient at their time of arrival and was brought with them throughout the course of their time at the clinic. The time of day was recorded when they began service with a provider, and when they departed the clinic. This allowed the appointment length for all patients who used this tracking tool to be computed.

As seen in Table 1 the data was recorded for walk-in and booked patients separately. The sample mean and standard deviation for booked patient service times were 11.03 minutes and 6.87 minutes, respectively; for walk-in patient service times, they were 11.80 minutes and 7.48 minutes.

Table 1: Patient Appointment Length Data

	Patient Appointment Length		
	Walk-In	Booked	Combined
Average	11.80	11.03	11.27
St. Dev	7.48	6.87	7.02
Sample Size	25	58	83

Due to the negligible difference between these two patient types, the service time data was combined such that a single service time distribution was used for all patients, as depicted by the histogram in Figure 17. The sample mean and standard deviation for this combined data is 11.27 minutes and 7.02 minutes.

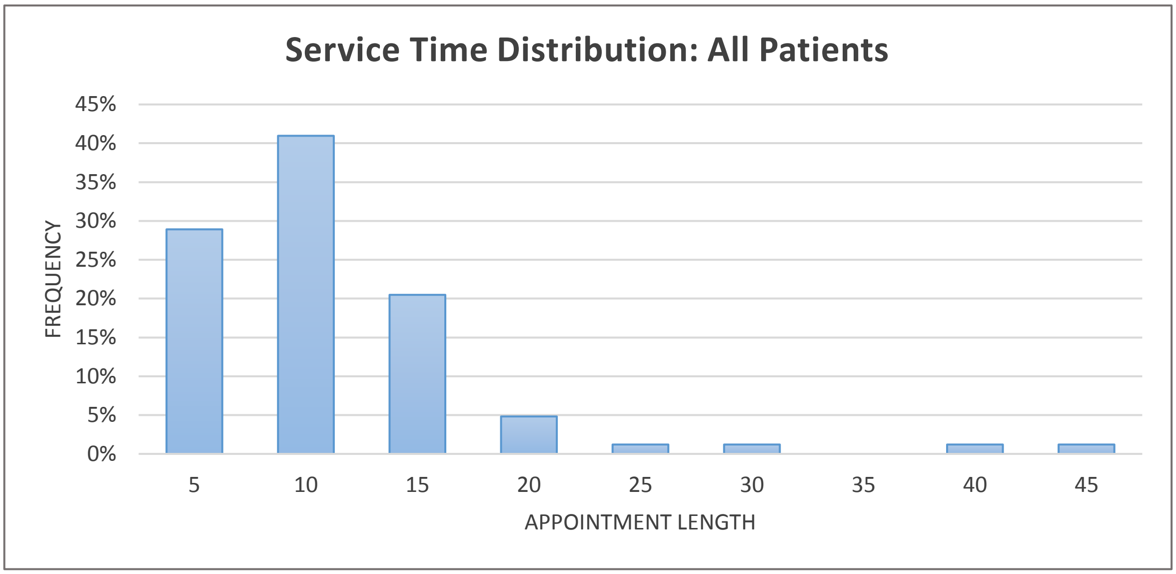

Figure 17: Patient Appointment Length Distribution

4.3. Case Study Input Parameters

This section describes the input parameters used in the model to characterize the CEC case study environment. The primary provider is scheduled to work an 8am-8pm clinic shift, however, there are two periods of unavailability during this time frame. From 12pm-1pm and 4:30pm-5pm the provider is unavailable due to meal breaks. The secondary provider is available to assist the primary provider in serving the walk-in demand from 9am-11am. Walk-in patients arrive to the system based on the distribution described in Chapter 4.2.1, which leads to an average daily volume of walk-in patients equal to 21.82. The patient appointment lengths are based on the distribution described in Chapter 4.2.2, which leads to an average appointment length equal to 11.27 minutes. The appointment buffer that is added to each of these appointment lengths is 5 minutes, representing the time that is required for things such as charting and notes following service. The performance metric that the clinic is most interested in is average patient waiting time, which should remain sufficiently small. In addition to average patient waiting time, the clinic is interested in maintaining a nominal value for average clinic overtime. The parameters gathered from this information are summarized in Table 2.

Table 2: Case Study Input Parameters

Parameter	**Values**
Primary Provider Availability	8am-12pm, 1pm-4:30pm, 5pm-8pm
Secondary Provider Availability	9am-11am
Walk-In Arrival Distribution	As described in Chapter 4.2.1 → Avg. daily volume = 21.82
Service Length Distribution	As described in Chapter 4.2.2 → Avg. length = 11.27 minutes
Buffer between appointments	5 minutes
Primary Metric of Interest	Average Patient Waiting Time
Secondary Metric of Interest	Average Clinic Overtime

4.4. Evaluation Approach

As indicated by the input parameters, the metrics used to ensure performance expectations are met include average patient waiting time, and average clinic overtime. It is expected that as the volume of booked patients increases, so too will the averages for patient waiting time and clinic overtime. Since it is desired to have low values for these metrics, there must be a criterion to indicate the maximum threshold values for these measures. Table 3 outlines the maximum for each value, indicating that average patient waiting time cannot exceed 30 minutes, and average clinic overtime cannot exceed 10 minutes. The aim is to schedule as many booked patients as possible while ensuring that the average values for patient waiting time and clinic overtime do not exceed these threshold values.

Table 3: Metric Max Value Criteria

Metric	Maximum Threshold Value
Average Patient Waiting Time	30 minutes
Average Clinic Overtime	10 minutes

Chapter 5 Results

The models presented in Chapter 3 will now be used to answer how many booked patients to schedule, and when to schedule them throughout the day. The model is run using parameters that represent the environmental factors unique to the Lillian Fraser Memorial Hospital's daytime CEC as introduced in 0.

In Chapter 5.1, a pseudo-optimal schedule template is determined for a variety of booked patient volumes using average patient waiting time as the objective function. How these schedule templates perform in terms of both average patient waiting time and average clinic overtime is observed. In Chapter 5.2, pseudo-optimal schedule templates for a variety of booked patient volumes are determined, now considering multiple objective functions.

5.1. Determine Appropriate Volume of Booked Patients

In this section, the effect that the volume of booked patients has on average patient waiting time and average clinic overtime is examined. The values of these metrics as computed by the simulation model will be compared to the values in Table 3.

The application of the simulation and SA models involve the following steps. For a given volume of booked patients, an initial schedule template is generated and 160 unique clinic days are replicated in the simulation model using this initial schedule template. From this simulation output, the performance metrics of interest are calculated and returned to the SA algorithm. The SA algorithm uses these metrics to compare with a newly generated neighboring schedule template, and accepting/rejecting this new schedule template when appropriate. This process is continued until a stopping criteria is met, and a pseudo-optimal schedule template has been determined. This is repeated for a range of booked patient volume values. In general, the run times for the case study below had a range of 2 minutes to 4 minutes.

This procedure is run many times with the volume of booked patients increasing for each run. In this section, the average patient waiting time is used for the objective function, as it is the primary metric of interest. Each run of the procedure finds a pseudo-optimal schedule template for the specified volume of booked patients for that model run. This pseudo-optimal schedule template indicates when throughout the day to schedule each booked patient such that the average patient waiting time is minimized. In addition to determining a pseudo-optimal schedule for each model run, the metrics of interest are also computed to evaluate the performance of the pseudo-optimal schedule.

Figure 18 illustrates the average utilization for the primary provider with 95% error bars. This metric is plotted as a function of booked patient volume, for values ranging from 0 to 30. It can be seen that when there are zero booked patients scheduled and the only system demand comes from walk-in patients, average utilization for the primary provider is 40%. Provider utilization increases as the volume of booked patients increases, approaching a near 100% utilization when an excess of 25 patients are booked.

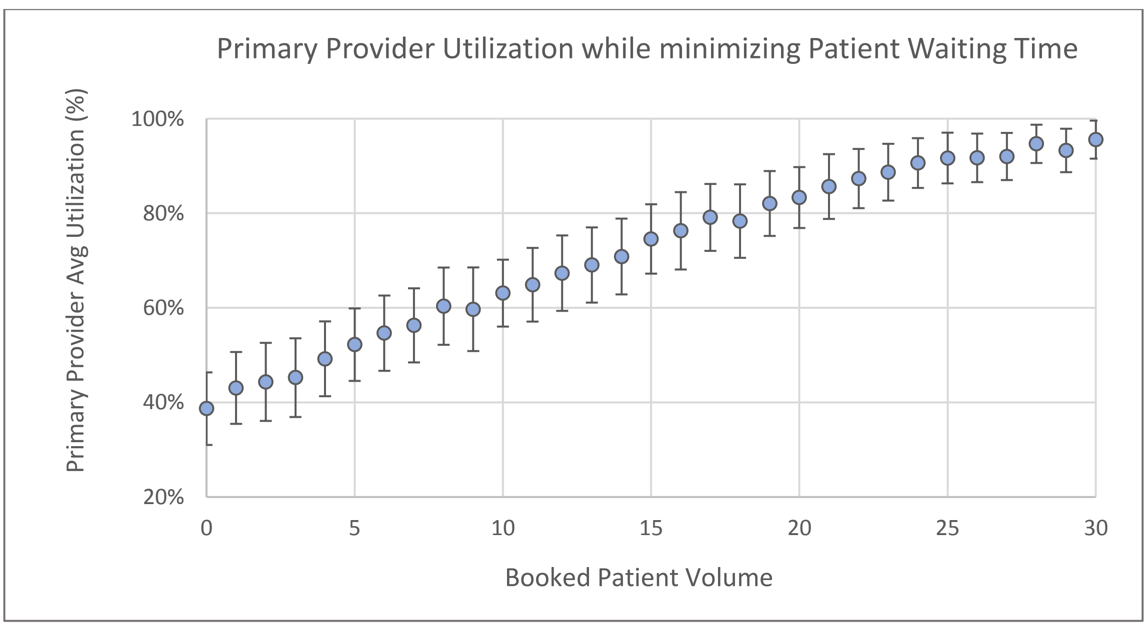

Figure 18: Impact of Booked Patient Volume on Primary Provider Utilization

Figure 19 details the average patient waiting time for booked patient volumes ranging from 0 to 30, with a hatched line representing the threshold value for this metric. As the volume of booked patients increases, the waiting time for patients increases exponentially.

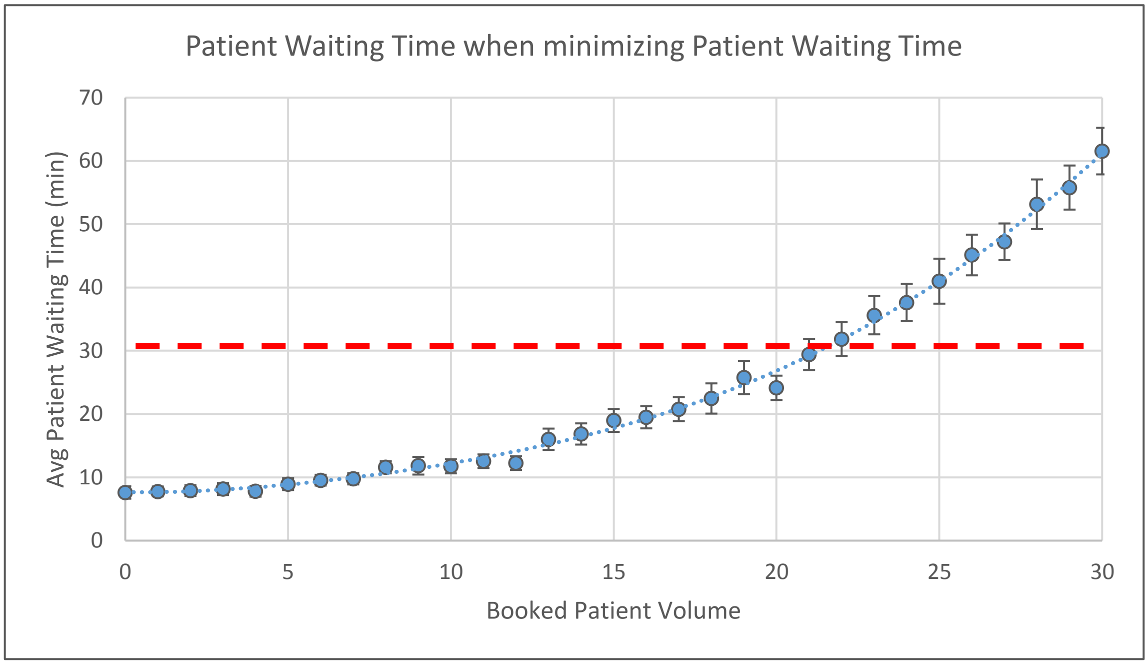

Figure 19: Impact of Booked Patient Volume on Patient Waiting Time

Considering that the maximum acceptable average waiting time for patients is 30 minutes, it can be seen that the maximum number of booked patients that adheres to this condition is 21. Before it is concluded that 21 is the maximum volume of booked patients that the clinic can handle, the clinic overtime threshold must also be considered.

Figure 20 illustrates the exponential relationship between booked patient volume and clinic overtime.

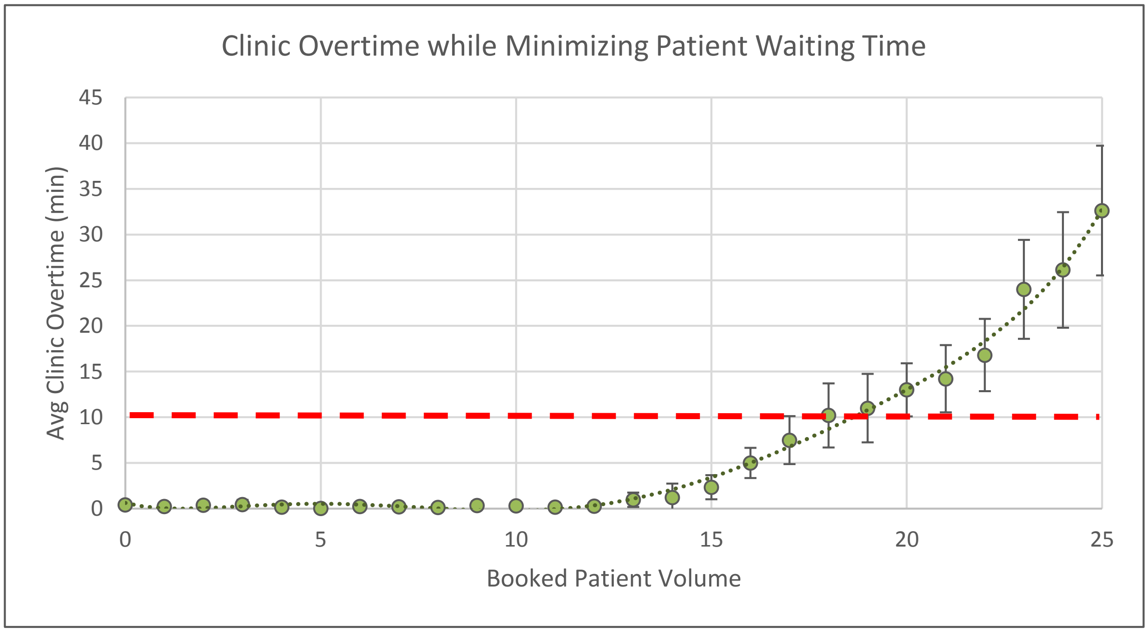

Figure 20: Impact of Booked Patient Volume on Clinic Overtime

For volumes of booked patients between 0 and 12 there is a negligible amount of overtime, indicating there is sufficient capacity for all booked patients and walk-in patients to be seen within the 8am-8pm time frame. As the volume of booked patients increases past 12, the average amount of overtime trends upward exponentially. Considering the average overtime threshold of 10 minutes, the corresponding volume of booked patients that does not exceed this threshold is 17.

Therefore, when minimizing average patient waiting time as the objective, a booked patient volume of 17 is the maximum allowable such that both thresholds are adhered to. Figure 21 and Figure 22 explore this further and highlight the slack in the patient waiting time threshold associated with scheduling 17 patients.

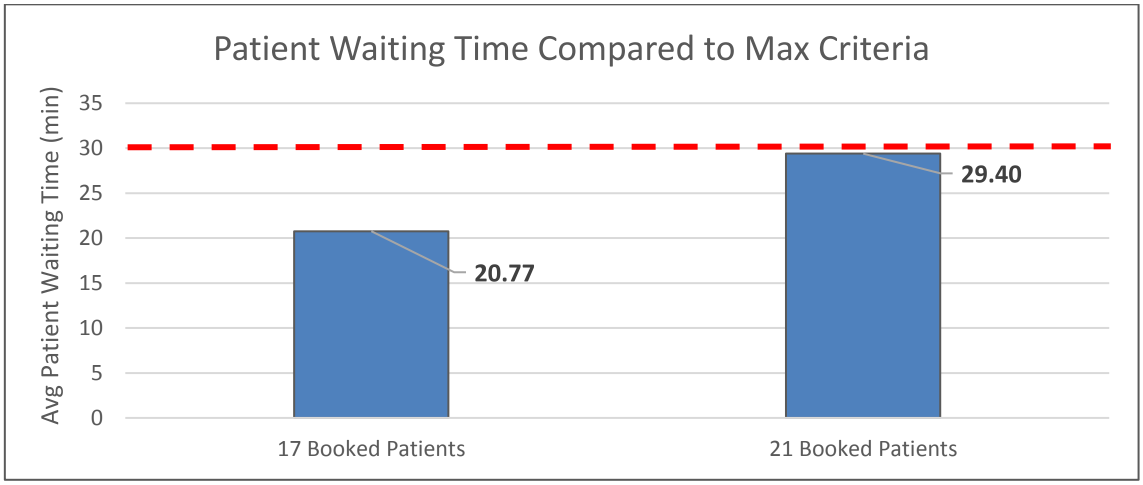

Figure 21: Patient Waiting Time Compared to Max Criteria

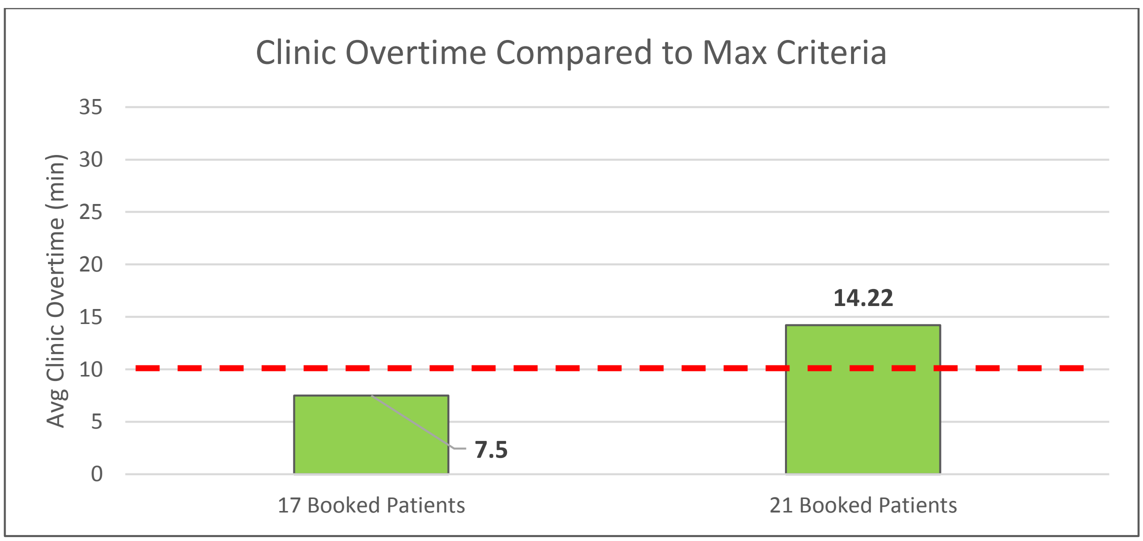

Figure 22: Clinic overtime at critical volumes

When scheduling 21 booked patients, the average patient waiting time is very close to meeting the threshold for average patient waiting time, but the average clinic overtime exceeds the threshold value. When scheduling 17 booked patients, average clinic overtime is very close to meeting the threshold value, but the average patient waiting time is much lower than the threshold value for this metric. In the following section, the volume of patients that can be booked is increased by aligning the metrics of interest as close to the threshold values as possible. This was accomplished by adjusting the objective function to include clinic overtime in addition to patient waiting time.

5.2. Balance Patient Waiting Time and Clinic Overtime

It will now be shown how the two metrics of interest can be balanced by aligning their performance as close to their threshold value as possible. This was accomplished by adjusting the objective function (OF) such that both clinic overtime and patient waiting time are considered. The aim in conducting this change was to allow for an increased volume of booked patients to be seen while still meeting both metric criteria.

The simulated annealing procedure was run for booked patient volumes of 0 to 21 for three differing objective functions, as shown below. This leads to 66 pseudo-optimal schedule templates and their associated metrics. These metrics are then used to determine which template allows the most booked patients to be scheduled while still adhering to both metric criteria.

$$OF1 = 1.0 * PatientWait + 0.0 * Overtime$$

$$OF2 = 0.5 * PatientWait + 0.5 * Overtime$$

$$OF3 = 0.0 * PatientWait + 1.0 * Overtime$$

Each objective function is a weighted sum of average patient waiting time and clinic overtime. $OF2$ gives them equal weighting, while $OF1$ and $OF3$ examine the extremes.

5.2.1. Average Patient Waiting Time with Multiple Objective Functions

Figure 23 plots the average patient waiting times computed from the SA procedure, using the three objective functions and a range of booked patient volumes.

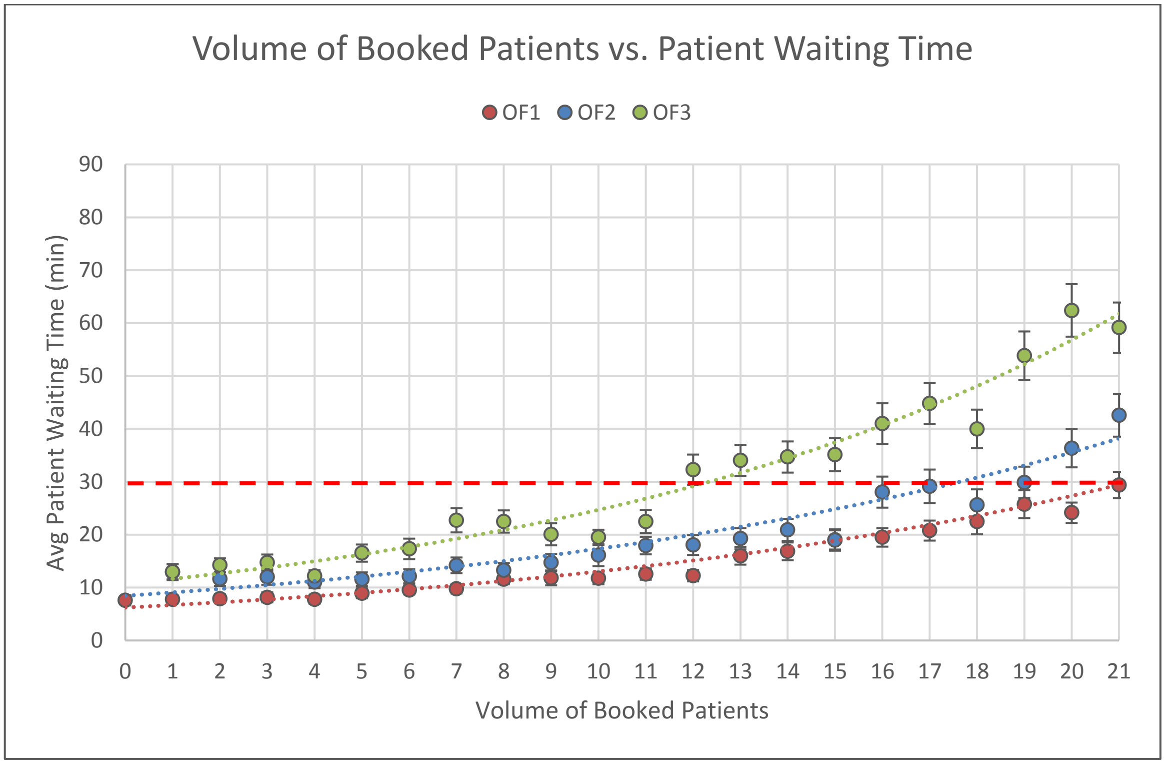

Figure 23: Patient Waiting Time as a function of Booked Volume and OF

As expected, when average patient waiting time is no longer the sole focus in the objective function, there is an increase in the average patient waiting time values as shown by the results of $OF2$ and $OF3$. When the objective function is solely minimizing clinic overtime ($OF3$) , the maximum volume of booked patients that can be seen while still meeting the patient waiting time criteria is 11.

However, when the objective function is an equally weighted sum of patient waiting time and clinic overtime ($OF2$), the critical number of booked patients based on average patient waiting time is now 19. This is greater than the previous critical value of 17, therefore, analysis can be continued to determine how clinic overtime is affected by these objective functions.

5.2.2. Average Clinic Overtime with Multiple Objective Functions

Figure 24 plots the average clinic overtime values computed from the SA procedure, using the three objective functions and a range of booked patient volumes.

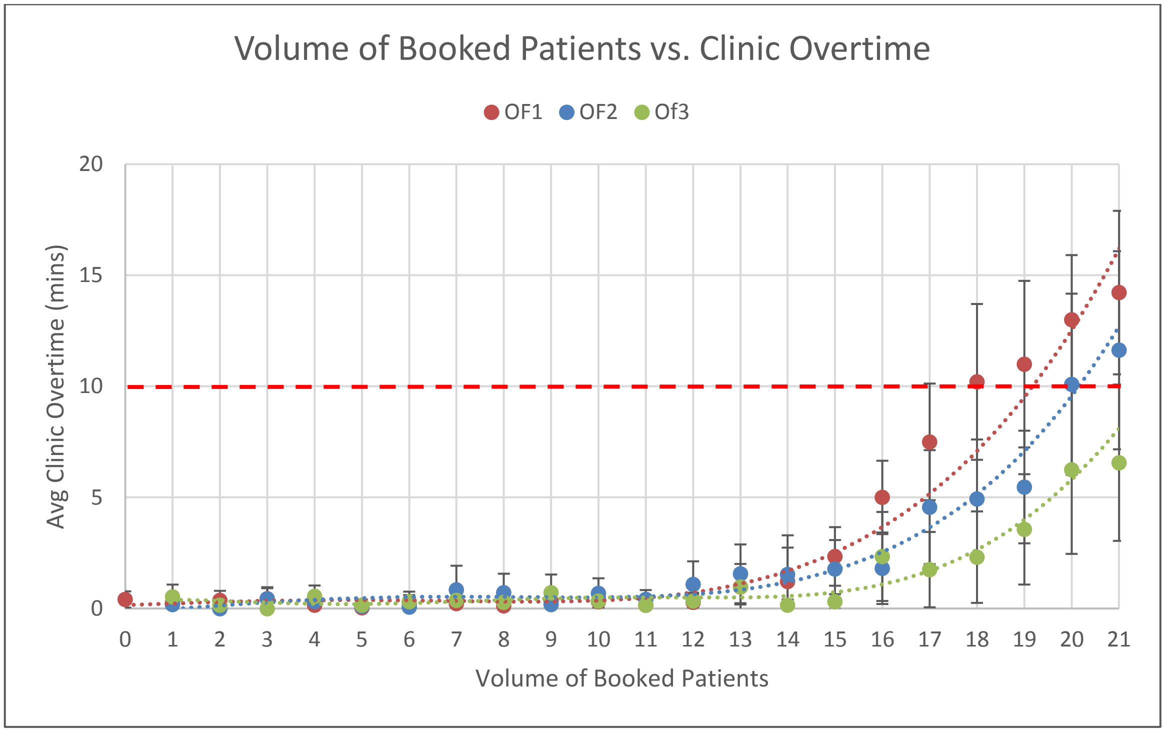

Figure 24: Clinic Overtime as a function of Booked Volume and OF

It can be seen that when using an equally weighted combination of patient waiting time and clinic overtime as the objective function ($OF2$), the critical number of booked patients is 19.

The critical value for the maximum number of booked patients to schedule based on the set criteria is now chosen as the minimum of the two critical values for patient waiting time and clinic overtime. In this instance, the critical volume of booked patients based on average patient waiting time and average clinic overtime are both 19. Therefore, a critical volume of 19 is chosen as an estimate of the suitable volume of booked patients to align with the metric criteria. Figure 25 and Figure 26 illustrate the balancing of the deficit between the average metric and its maximum value criteria. This was accomplished by adjusting the objective function to include clinic overtime, such that average patient waiting time increased closer to the threshold value, and average clinic overtime remained fairly stagnant with the addition of booked patients.

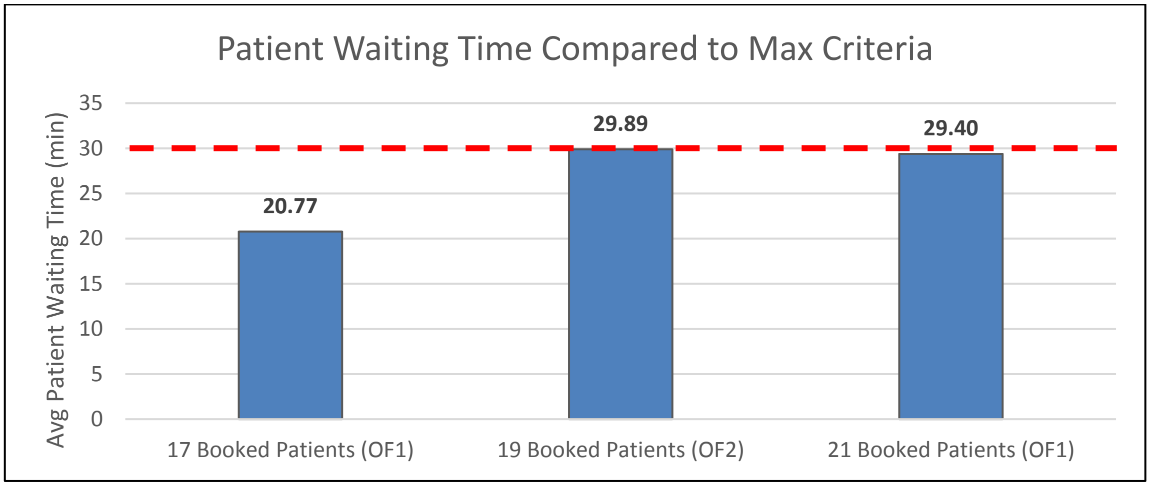

Figure 25: Critical volumes based on patient waiting time vs. threshold

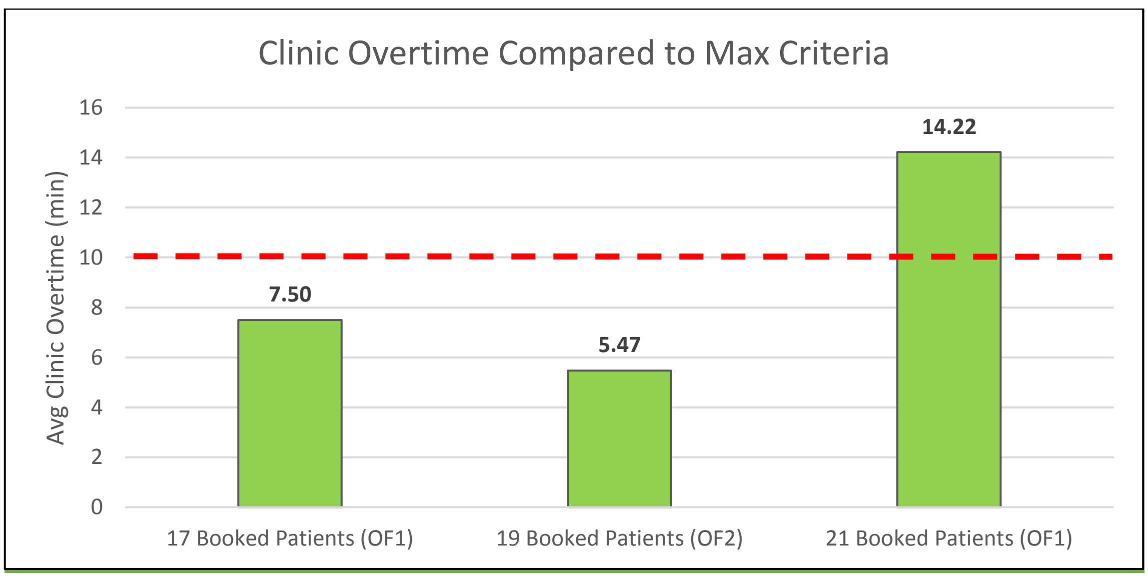

Figure 26: Critical volumes based on clinic overtime vs. threshold

Table 4 summarizes the results from the analysis using these three objective functions. The critical volume of booked patients based on each metric for the three objective functions tested are detailed. It is clear, that 19 is the maximum volume of patients that can be scheduled such that both thresholds are adhered to.

Table 4: Critical volumes based on metric and OF

		OF1	OF2	OF3
Critical Booked Patient Volume Based on:	Patient Waiting Time	21	19	11
	Clinic Overtime	17	19	21
	Both	17	**19**	11

Finally, to explore the sensitivity of the objective function weights, the analysis used to generate Table 4 is repeated with varying α and β used in the objective function $OF = \alpha * PatientWait + \beta * Overtime$. The results of this extended analysis are shown in Table 5, with results indicating 19 booked patients is the most that can be scheduled to meet the metric criteria. This occurs when the objective function is an equally weighted sum of the two metrics of interest.

Table 5: Extended objective function analysis

	α	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
	β	1	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1	0
Clinic Booked Patient Volume Based on:	**Patient Waiting Time**	11	15	18	17	18	19	19	18	19	20	21
	Clinic Overtime	21	20	20	20	19	19	18	18	18	17	17
	Both	11	15	18	17	18	**19**	18	18	18	17	17

5.3. Review of Pseudo-Optimal Schedule Template

Recall that each time the SA procedure is run, a pseudo-optimal schedule template and computed values for the metrics of interest are obtained as an output. The metrics were used in the previous sections to determine what a suitable volume of booked patients for the clinic should be. Upon determining that 19 is a reasonable volume of patients to book in a day, the associated pseudo-optimal schedule template can be reviewed. This pseudo-optimal schedule template is shown in Figure 27. In addition to the booked patient schedule, this figure also plots the average number of walk-in patients in the system throughout the day.

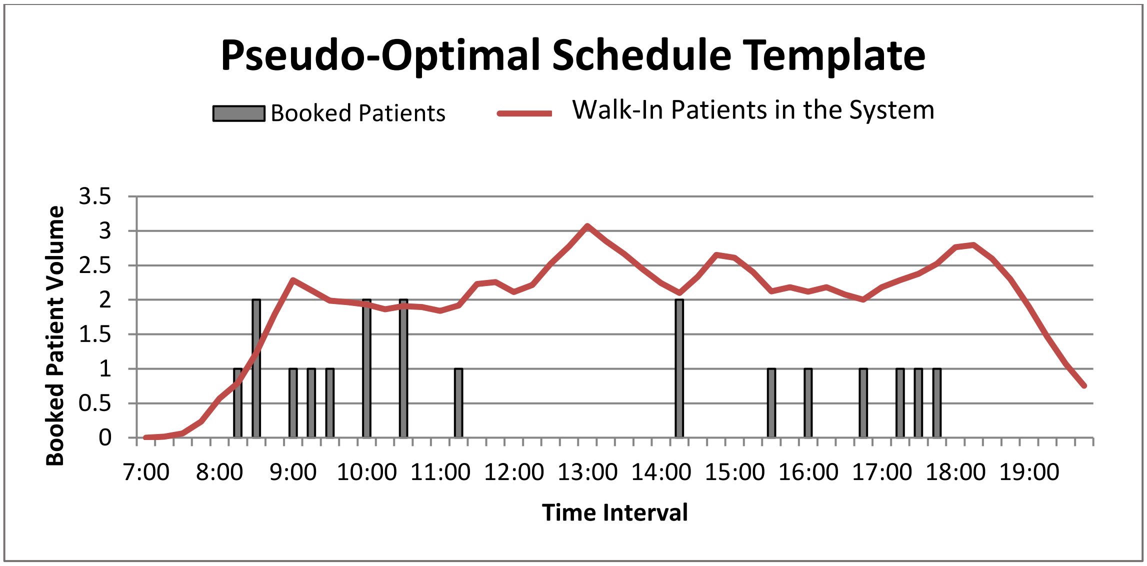

Figure 27: Pseudo-optimal schedule template

The first 3 patients are given an appointment time between 8am and 9am. At this point in the day, the backlog of walk-in demand is still increasing such that the primary provider will be able to see booked patients without significantly impacting waiting time for walk-in patients. The next 8 patients are booked between 9am and 11am. This too is a logical period to place booked patients since the secondary provider is available to assist with walk-in demand during this time.

As expected there are no booked patients scheduled between 12pm and 1pm, which is the period of time when the primary provider takes lunch. During this period, the average volume of walk-in patients in the system is steadily increasing because no patients are being seen during this time.

There are no patients booked between 1pm and 2:15pm, which is the period of time directly following the provider lunch break. This is allowing for a portion of the patient demand built up over lunch to be served prior to more booked patients being added to the system.

The remaining 8 booked patients are dispersed throughout the afternoon, with the final booked patient occurring at 5:30pm. This allows for the patient demand remaining at the end of the day to be served in less than 10 minutes on average, indicating the maximum average overtime criteria has been met.

5.4. Secondary Provider Availability Case Study

The results to this point have shown how the appropriate volume of booked patients and an associated pseudo-optimal schedule can be determined. This was under the assumption that a secondary provider was necessary and scheduled to assist with walk-in demand between 9am-11am. The aim of this section is to demonstrate how the model can be used to determine how best to utilize the secondary provider, i.e. to determine the best period of time throughout the day to make them available. Recall that the secondary provider is only able to see walk-in patients, not booked patients, and will be the priority provider for walk-in patients while they are available. For simplicity, only the patient waiting time metric is evaluated in this demonstration. The input parameters used in this testing can be seen in Table 6.

Table 6: Secondary Provider Analysis Parameters

Parameter	Values
Primary Provider Availability	8am-12pm, 1pm-4:30pm, 5pm-8pm
Walk-In Arrival Distribution	As described in Chapter 4.2.1 → Avg. daily volume = 20.82 patients
Service Length Distribution	As described in Chapter 4.2.2 → Avg. length = 11.27 minutes
Buffer between appointments	5 minutes
Metric of Interest	Average Patient Waiting Time

5.4.1. Determine if a secondary provider is necessary

To determine whether it is necessary to use a secondary provider to assist with walk-in demand, it is observed how patient waiting time is affected when there is no secondary provider availability. This will be compared to the previous scenario, where the secondary provider was available for two hours from 9am to 11am.

Figure 28 illustrates that as expected, there is significant improvement in patient waiting time when a secondary provider is available for two hours in the morning from 9am to 11pm. The difference in patient waiting time becomes increasingly pronounced as the volume of booked patients' increases.

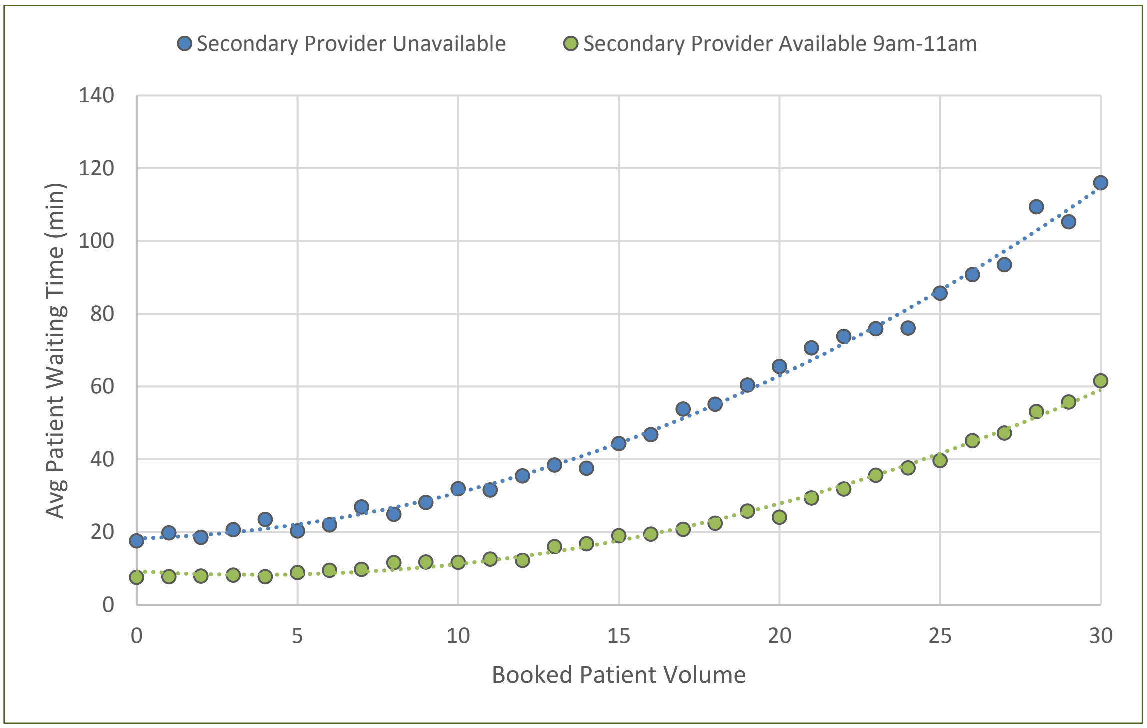

Figure 28: Impact of secondary provider availability on patient waiting time

These results indicate that there is a significant benefit to having a secondary provider available to assist with the demand of walk-in patients.

5.4.2. Placement of secondary provider availability

Having determined that using the services of a secondary provider is beneficial to the system, it must also be determined where to strategically align the availability of the secondary provider such that the most benefit can be seen. Assume that the amount of time the secondary provider has been allotted to be available is two hours. As shown in Figure 29, the highest walk-in patient demand occurs during the morning session of the clinic. Therefore, analysis will be conducted to examine the best 2- hour period in the morning session to make the secondary provider available to assist with walk-in demand.

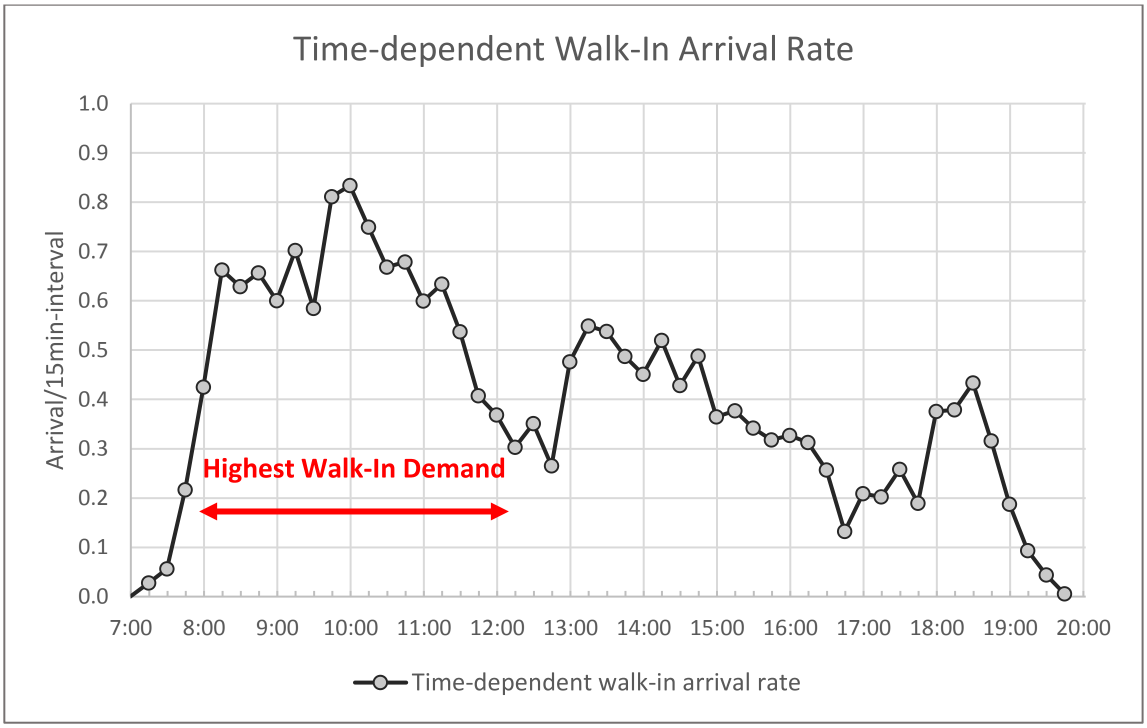

Figure 29: Time Dependent Walk-In Arrival Rates

Clinic performance is evaluated at three distinct 2-hour periods of availability in the morning session (8am-10am, 9am-11am, and 10am-12pm). Testing is completed to determine the clinic performance with an objective of minimizing patient waiting time.

As seen in Figure 30, the relative difference in performance is quite small for these varying periods of secondary provider availability. The difference in performance increases as the volume of booked patients increases. Therefore, depending on the volume of booked patients desired, the recommendation for which availability period to implement may vary.

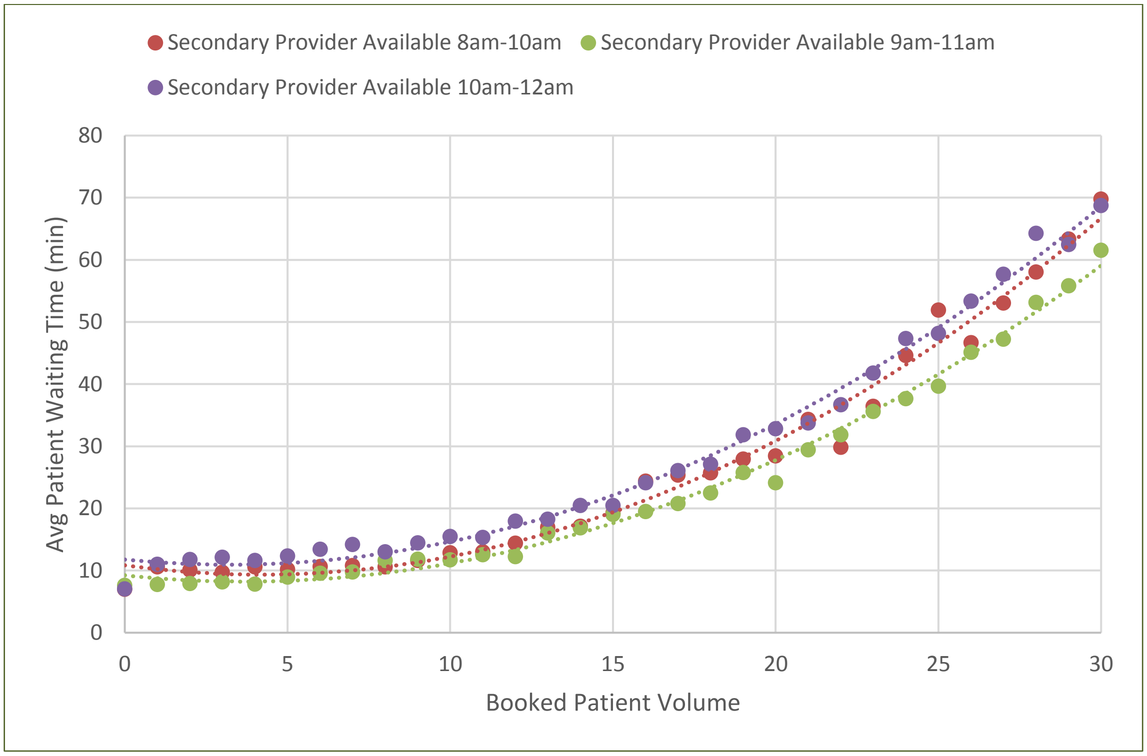

Figure 30: Impact of secondary provider availability placement on waiting time

The availability period between 9am and 11am is trending as the best availability period to use for all volumes of booked patients. However, the superiority of this time period is quite small for low volumes of booked patients. Therefore, when the volume of booked patients is low, it would be recommended to choose the availability period that is most convenient for the clinic and secondary provider. As the volume of booked patients increases, and the difference in performance becomes more pronounced, the 9am to 11am availability period would see benefits in the clinic operations in regard to average patient waiting time.

5.4.3. Additional Secondary Provider Availability

It has been determined that secondary provider availability is beneficial, and that the best placement for two hours of availability is between 9am and 11am. Additional testing is now conducted to observe the impact of adding an additional hour of secondary provider availability in the afternoon clinic session. Testing was conducted to observe the effect on average patient waiting time when an additional hour of availability is placed between 2pm and 3pm.

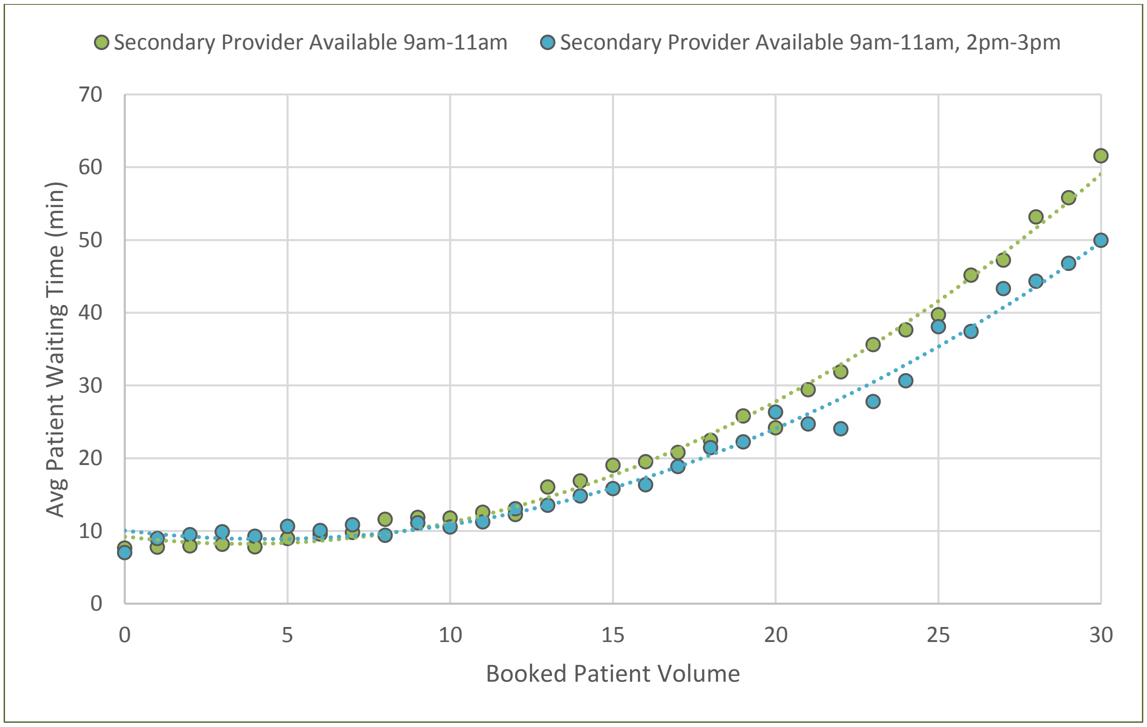

Figure 31: Impact of additional secondary provider availability on waiting time

Figure 31 shows that the additional availability would not lead to any improvement in patient waiting time when there are less than approximately 15 booked patients. With greater than 15 booked patients, the additional secondary provider availability becomes increasingly beneficial to the system as the volume of booked patients increases. Previous analysis showed that the system can handle approximately 19 booked patients. When the booked patient volume is 19 the improvement on patient waiting time with the additional availability does not appear significant enough to add availability in the afternoon session.

Chapter 6 Tool

This chapter introduces the tool that was developed for stakeholders to conduct independent analysis. It shows the development of the tool as well as the capabilities and functionality. The interface and steps taken to evaluate the current state are shown first, followed by an explanation of the different adjustments that can be made and evaluated. The tool was developed using Tkinter, which is the standard graphical user interface package for Python.

Figure 32 shows the landing page for the evaluation tool, with links to the four main actions that can be completed. These involve three varying methods of evaluation and access to the raw inputs used in the analyses, which are detailed in the following subsections.

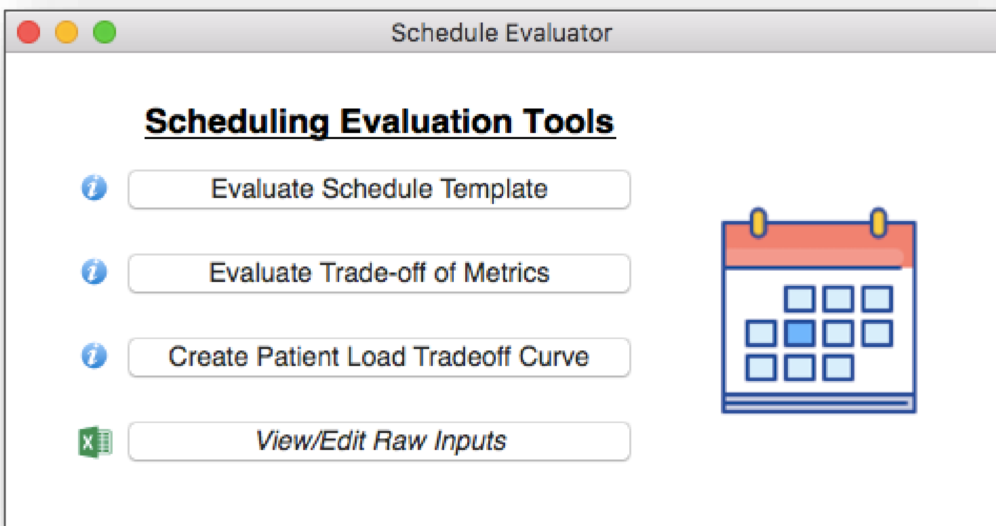

Figure 32: Tool Landing Page

For information on the three scheduling evaluation methods, clicking the information button (white “i" in the blue circle) to the left of the link will produce a pop-up window that provides a brief description of the methods and the inputs required. An example of this is shown for the “Evaluate Schedule Template” evaluation option in Figure 33.

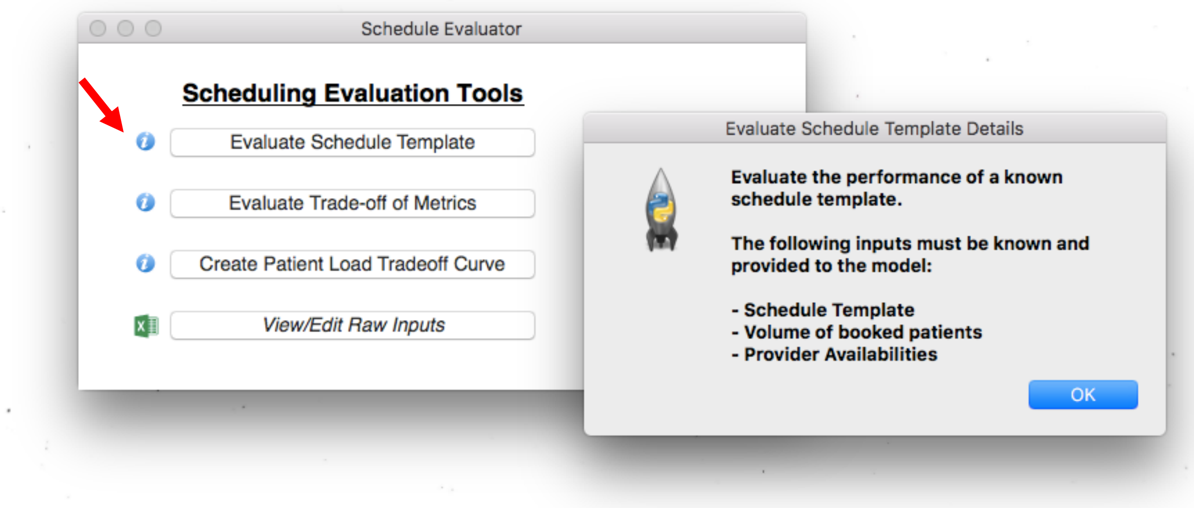

Figure 33: Evaluation Method Information Example

6.1. Evaluate Schedule Template

When a stakeholder would like to evaluate the performance of a known schedule template, they begin by clicking the "Evaluate Schedule Template" button. The required inputs include the volume of booked patients, when they are scheduled throughout the day, and the periods of availability for the primary and secondary provider. The input interface can be seen in Figure 34.

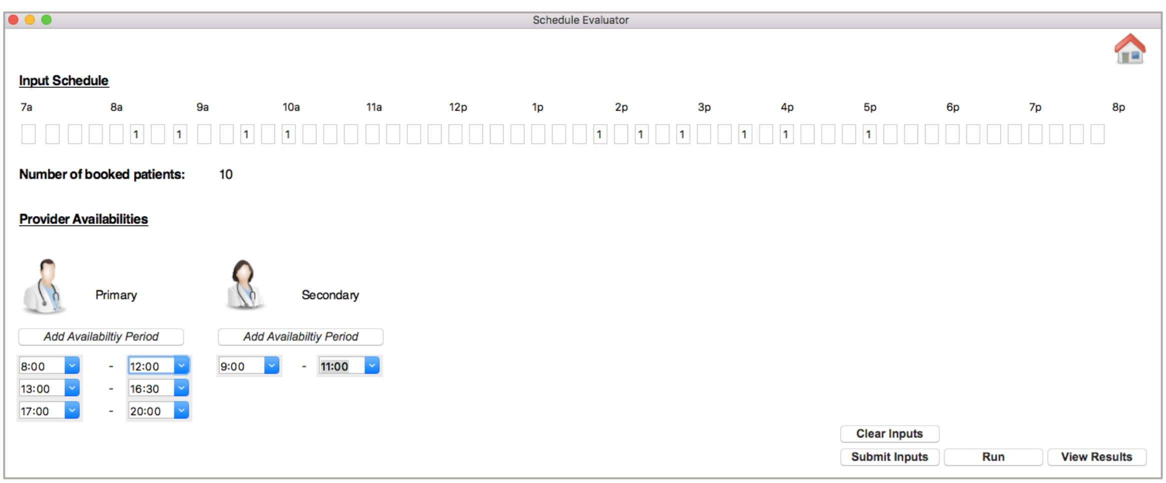

Figure 34: Schedule Template Evaluation Input Interface

Once the inputs are submitted and the "Run" button clicked, the "View Results" button will appear. Clicking this button leads the user to the screen shown in Figure 35, which displays the performance metrics averaged across all replicated days.

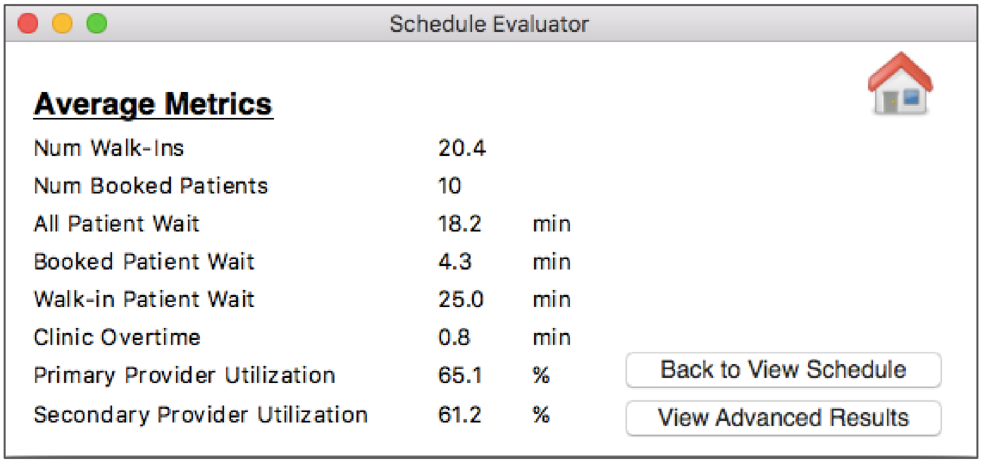

Figure 35: Schedule Evaluation Average Performance Metrics

These results provide average measures for how long patients are waiting in the system prior to being seen by a provider. This is further broken into average waiting times for booked patients and walk-in patients. Since booked patients have priority in the system, they have a much shorter average waiting time. Average overtime is provided to indicate the average amount of time that the primary provider is required to stay past the time he or she has been scheduled to work. The average amount of overtime in this example is quite low. Finally, average measures of utilization for the primary and secondary providers are shown. These represent the percentage of available working time used to see patients.

From this screen, the user can return to the main menu using the home icon in the top right corner, return to schedule inputs by clicking the "Back to View Schedule" button, or see additional metric information by clicking the "View Advanced Results" button. By clicking the "View Advanced Results" button, an Excel document (see Figure 36) that contains distributions for patient waiting time, booked patient waiting time, walk-in patient waiting time, clinic overtime, clinic idle time, and primary provider utilization will be opened. These results can be used to determine what percentage of patients have a waiting time less than a predetermined number of minutes, and other statistics beyond the means.

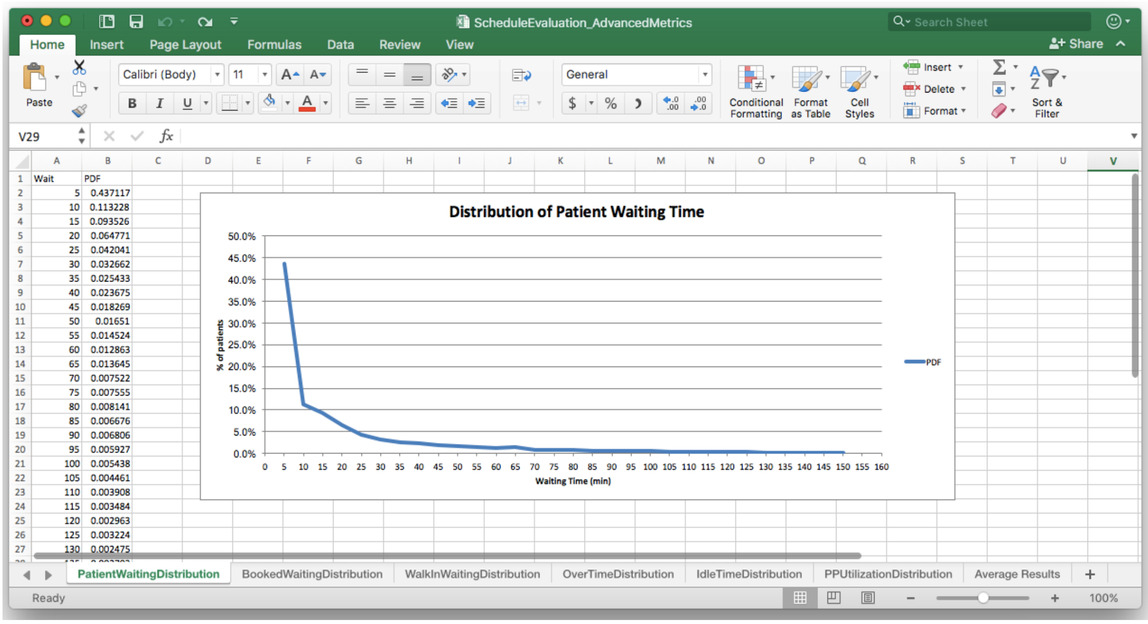

Figure 36: Advanced Metrics for Schedule Evaluation

6.2. Evaluate Trade-Off of Metrics

The second method of evaluation allows for the trade-off between two clinic metrics to be observed. This is accomplished by conducting multiple runs of the SA procedure with an objective function that is a weighted combination of the two metrics of interest. The scaling factors for these metrics are adjusted for each run. Let the scaling factor for the first metric of interest be denoted by α, and the scaling factor for the second metric of interest be denoted by β. The model allows the user to select two of the following three metrics for relational comparison:

- Patient Waiting Time
- Clinic Overtime
- Primary Provider Utilization

Using patient waiting time and clinic overtime as the metrics of interest, the objective function for the SA procedure would looks as follows:

$$\boldsymbol{Minimize}: \quad \boldsymbol{Cost = \alpha * PatientWaitingTime + \beta * ClinicOvertime}$$

The tool then runs the SA procedure 11 times, starting with $\alpha = 1$ and $\beta = 0$. After each run, α and β are decremented and incremented respectively by 0.1 until $\alpha = 0$ and $\beta = 1$.

Table 7: Simulated Annealing Procedure Schedule

Run #	$\boldsymbol{\alpha}$	$\boldsymbol{\beta}$
1	1.0	0.0
2	0.9	0.1
3	0.8	0.2
4	0.7	0.3
5	0.6	0.4
6	0.5	0.5
7	0.4	0.6
8	0.3	0.7
9	0.2	0.8
10	0.1	0.9
11	0.0	1.0

This means that in the first run of the procedure, the model is searching for a pseudo-optimal schedule that minimizes only patient waiting time with no concern for the adverse effects on clinic overtime. As the runs progress, the model becomes increasingly concerned with clinic overtime in addition to patient waiting time. This is true until the final run, when the model is solely minimizing clinic overtime with no concern for patient waiting time. The inputs required for this analysis includes the number of booked patients, two metrics of interest, and provider availabilities. The input interface is shown in Figure 37.

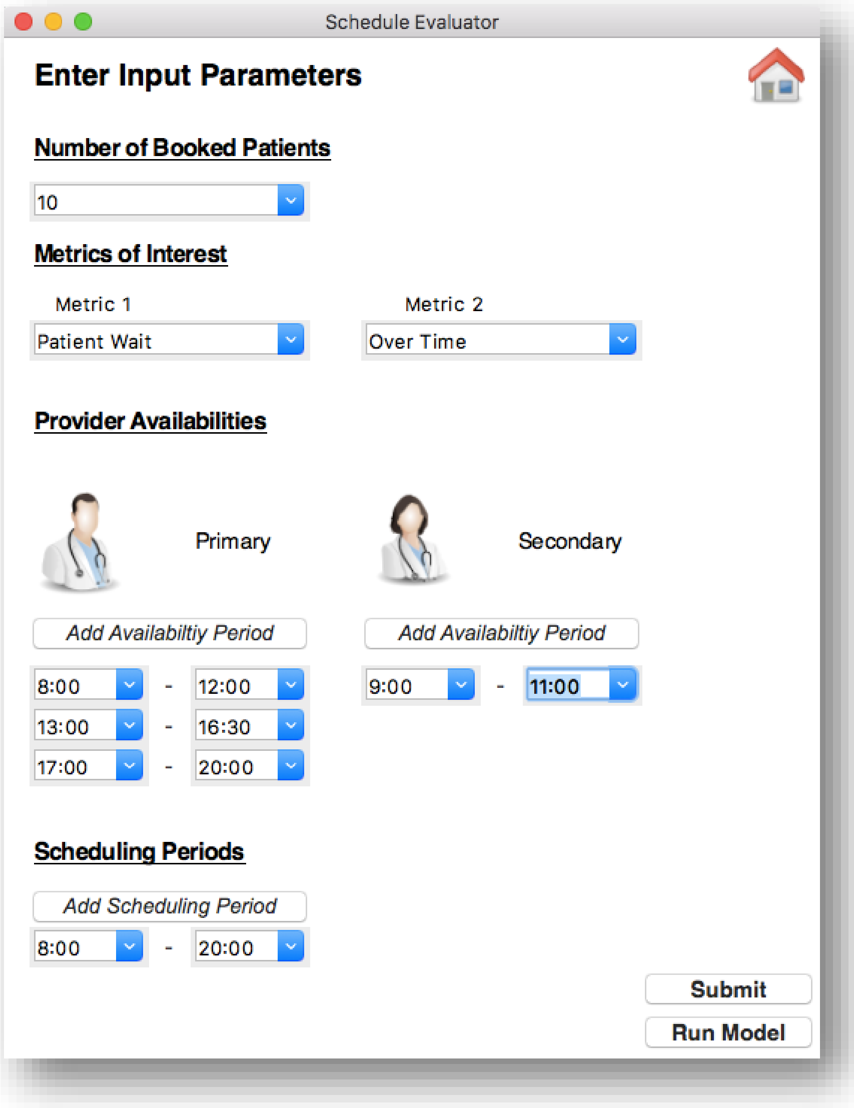

Figure 37: Pareto Frontier Interface

The output for this analysis includes the average metrics that are used to evaluate performance, and the pseudo-optimal schedules that are developed on each of the eleven runs of the simulated annealing procedure. A trade-off curve is plotted that compares the two metrics of interest on each of the runs. An example of this can be seen in Figure 38, where the trade-off between patient waiting time and clinic overtime is being depicted. The lowest values for each of these metrics occur when the simulated annealing procedure is run for each of the metrics individually. These values are obtained from run number 1 and 11 as shown in Table 7, and are noted in Figure 38 as extremes.

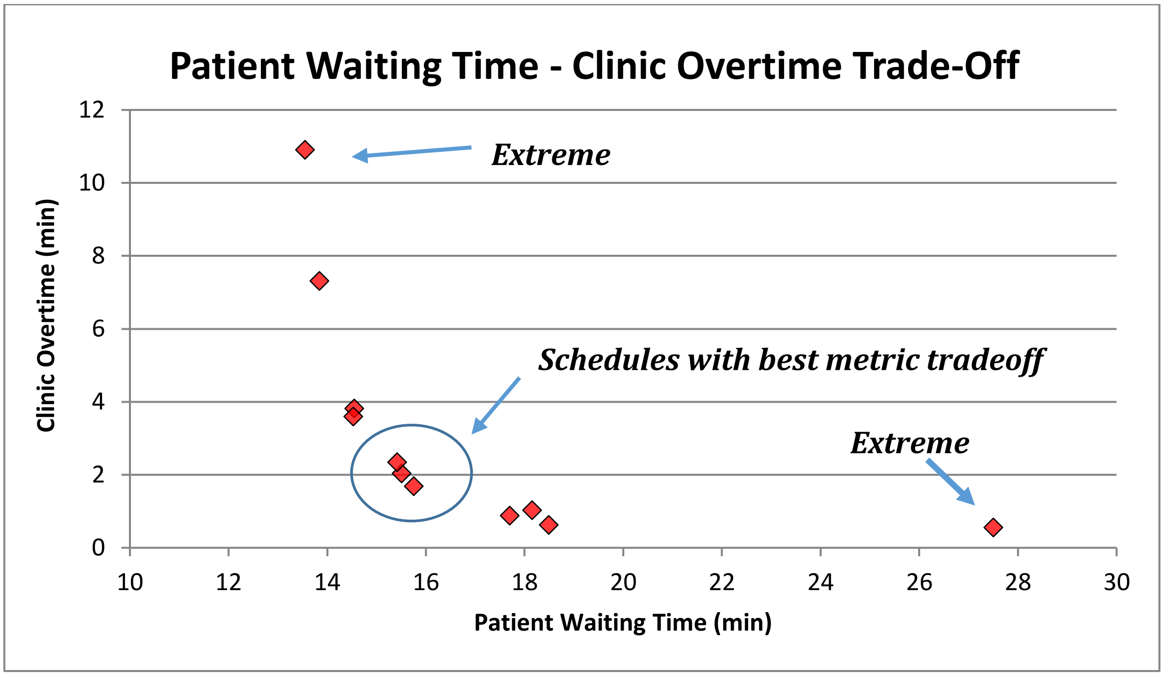

Figure 38: Metric Comparison Trade-Off Curve

This type of analysis is most useful when the volume of booked patients the clinic would like to schedule is known, however, the appropriate objective function is not. Developing pseudo-optimal schedules that minimize varying trade-offs between the two metrics of interest allow for decision makers to select the one that has the most desirable trade-off. In the current example, decision makers may decide that the cluster of 3 schedules, as indicated in Figure 38, provide the best trade-off of patient waiting time and clinic overtime, and will be further reviewed for selection.

6.3. Create Patient Load Trade-off Curve

The final evaluation method allows the user to evaluate the impact that the volume of booked patients has on the system. This is accomplished by running the SA procedure multiple times for a specified range of booked patient loads. This evaluation method was used in Chapter 5 to assist in determining an appropriate volume of booked patients that the system can handle. The required inputs for this method of evaluation includes the range of booked patient volumes to be analyzed, the objective function to be minimized during the SA procedure, and the provider availabilities. The objective function can be a weighted combination of one or two of the following metrics of interests:

- Patient Waiting Time
- Clinic Overtime
- Primary Provider Utilization

The input interface is shown in Figure 39.

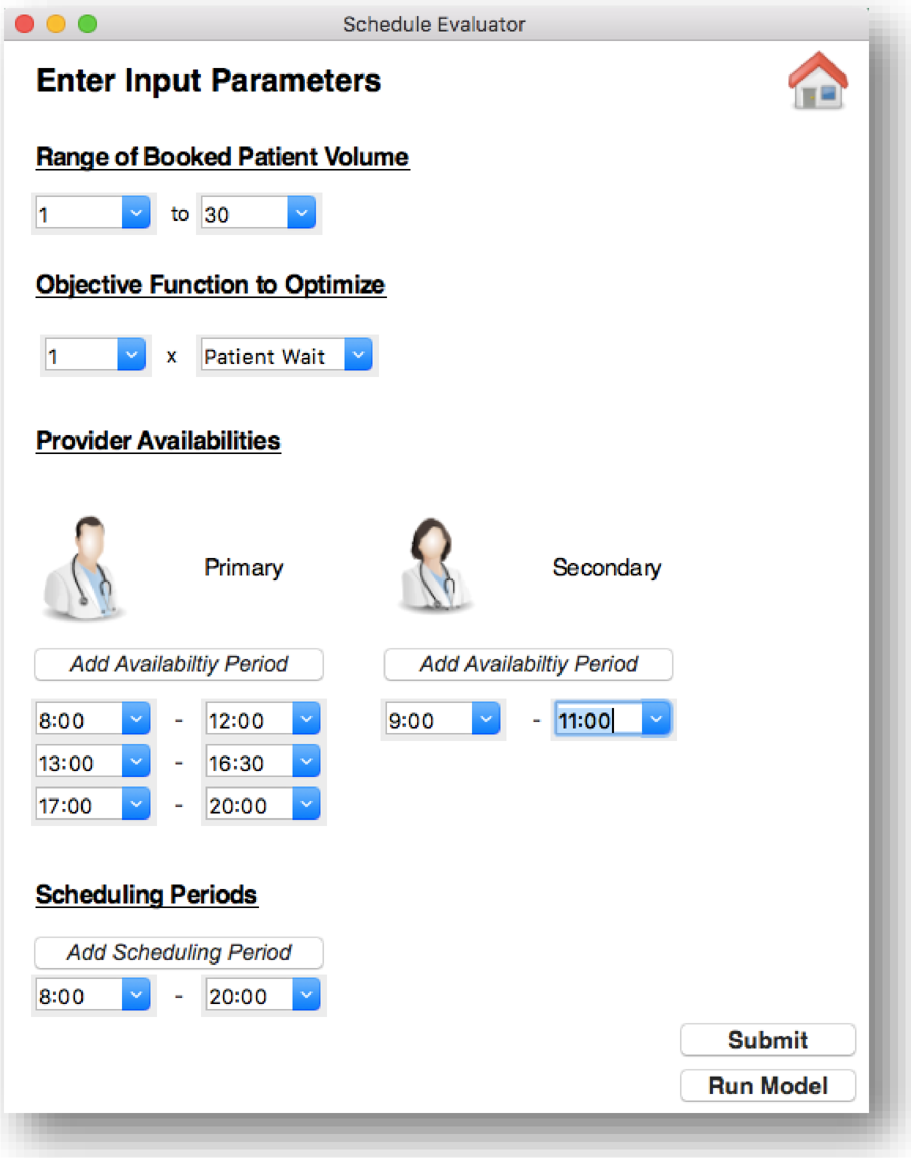

Figure 39: Patient Load Trade off Input Interface

Average clinic metrics including patient waiting time, clinic overtime, and provider utilization are calculated for each run of the simulated annealing procedure. Plots such as those seen in Figure 40, Figure 41, and Figure 42 are then produced. These allow the impact that patient volume has on each of these metrics to be observed.

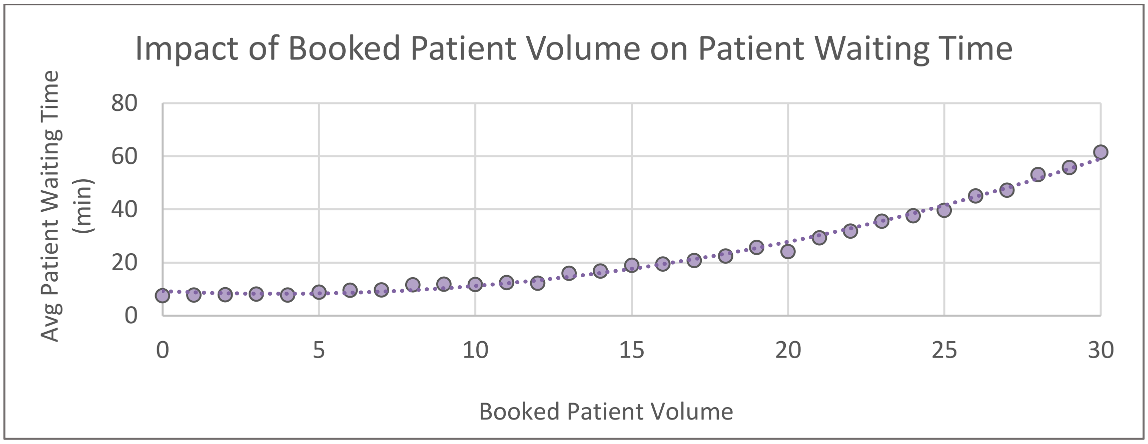

Figure 40: Impact of Booked Patient Volume on Patient Waiting Time

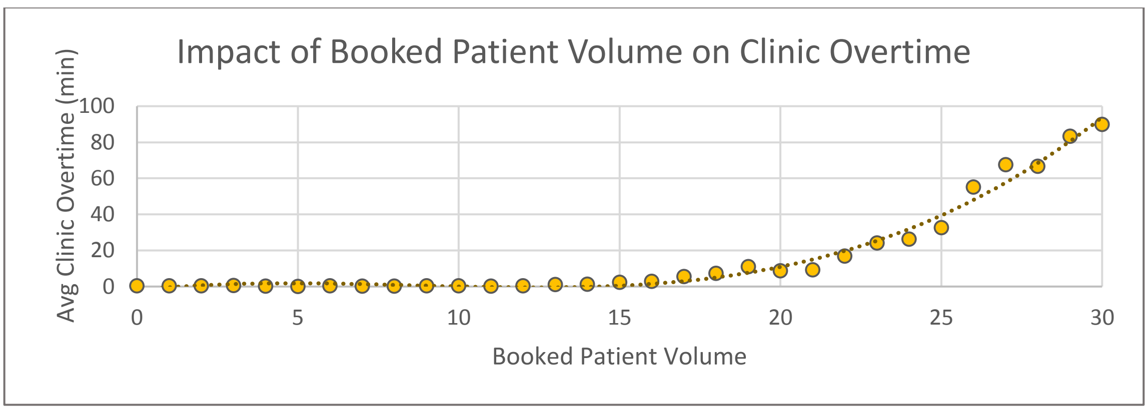

Figure 41: Impact of Booked Patient Volume on Clinic Overtime

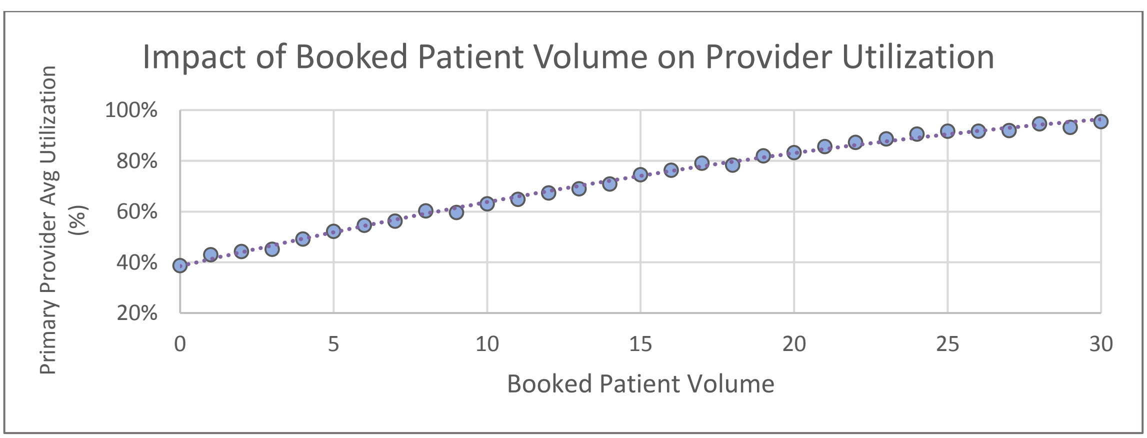

Figure 42: Impact of Booked Patient Volume on Provider Utilization

6.4. View Edit Raw inputs

The final capability of the user tool is to view and edit the raw data inputs required for the model. This includes the distributions for walk-in patient arrivals and appointment lengths, as detailed in Chapter 4.2. When a stakeholder would like to access the raw inputs, they begin by clicking the "View/Edit Raw Inputs" button on the tool landing page. This will open a page that indicates the raw inputs available for access. By clicking the Excel icon next to the indicated distribution as shown in Figure 43, an Excel file will be opened with the raw data for that distribution.

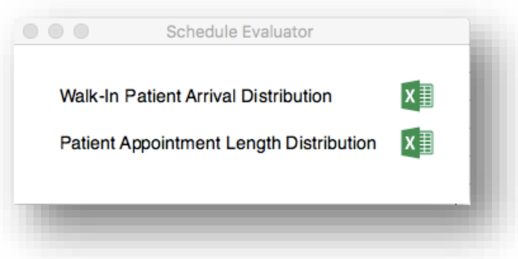

Figure 43: Raw Input Selection

The Excel sheet that contains the walk-in arrival rate data includes the time-dependent arrival rates, a graph that depicts the arrival pattern across the day, and a brief description of the data. This can be seen in Figure 44.

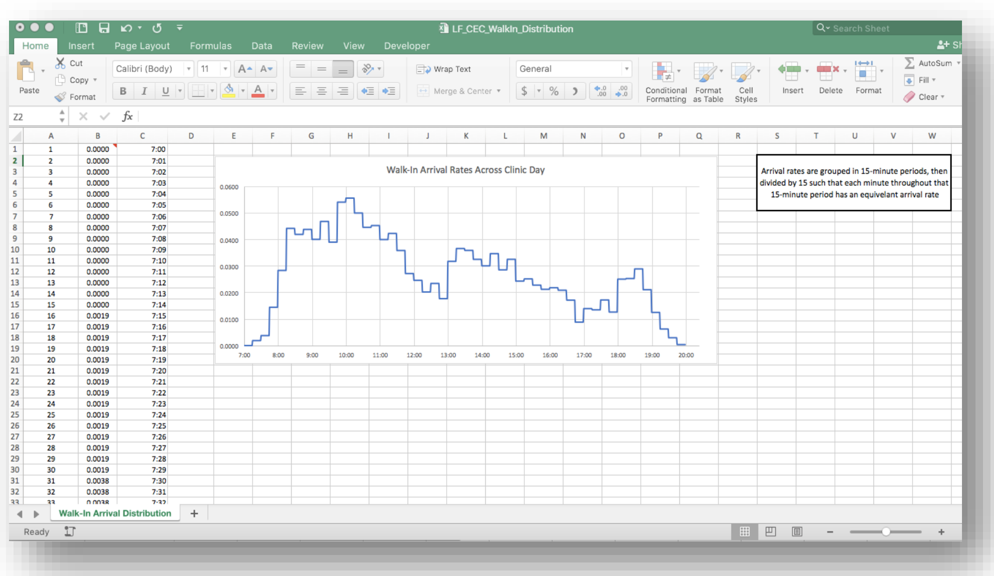

Figure 44: Walk-In Arrival Distribution Raw Data

The Excel sheet that contains the appointment length distribution data includes the probability of each appointment length, the appointment buffer, histograms that depict the probabilities, and a brief description of the data.

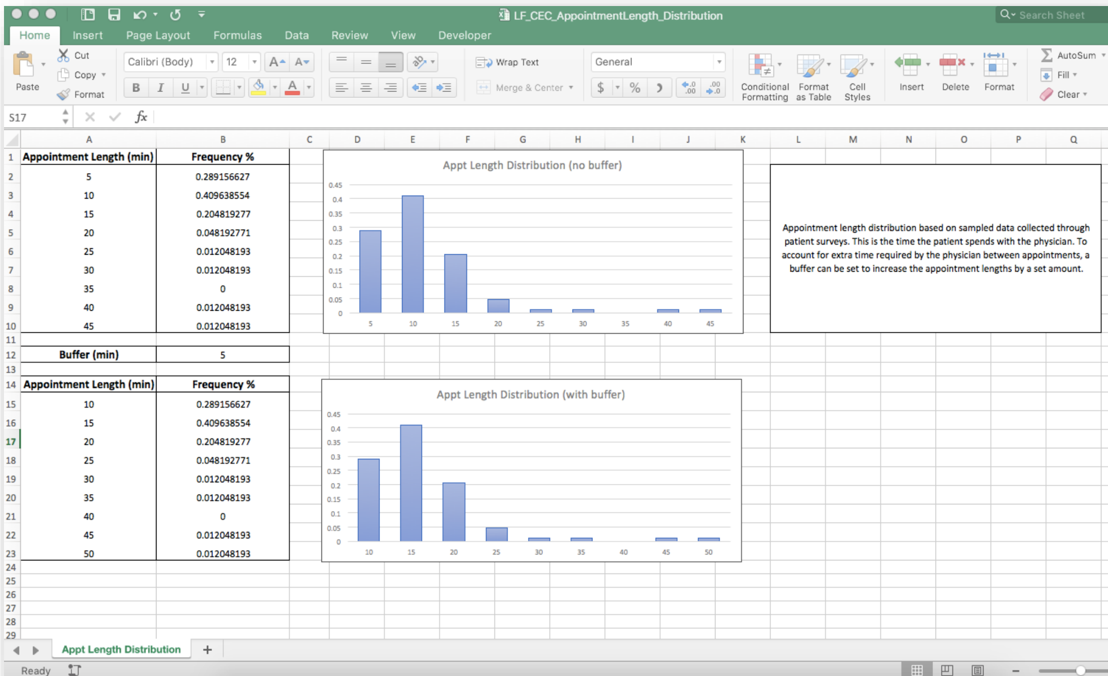

Figure 45: Appointment Length Distribution Raw Data

The data contained in these Excel sheets are clinic specific for the Lillian Fraser Memorial Hospital CEC. If the volume of walk-in patients arriving to the clinic, or the amount of time the provider takes with the patients' changes, this data may be adjusted to accurately reflect the clinic setting in its current state. Additionally, if evaluation was desired for a different clinic setting, data could be collected and input to these Excel sheets for analysis to be conducted.

Chapter 7 Conclusion

This research developed a method to determine the volume and time to schedule booked patients in an outpatient clinic with walk-in patients. This was accomplished using a simulation model of the clinic to compute estimates of average patient waiting time, clinic overtime, and provider utilization. In addition to evaluating a specified schedule template, simulated annealing was implemented to determine a pseudo-optimal schedule template.

One of the major benefits of the methodology is the ability to gain clinic specific results. Clinic specific factors such as non-stationary walk-in patient arrival patterns and appointment length distributions vary from clinic to clinic. Therefore, the approach allows users to input raw data specific to their clinics. The clinic may also have different metrics of interest, and threshold values for these metrics that must be met for a schedule template to be accepted. These metrics can be specified such that the clinic is developing a pseudo-optimal schedule template that meets their specific requirements.

To accomplish this, clinics must specify the walk-in patient arrival rates throughout the day, and the patient appointment length distribution for their clinic. They must also provide the clinic metrics that they are interested in, and threshold values for each which should not be exceeded. These are the constraints used to ensure suitable schedule templates are developed. Once these parameters are collected and specified in the model, testing and results can be gathered to determine a pseudo-optimal schedule template tailored to the specific clinic.

From the case study and analyses, a number of general rules of thumb have been determined to support clinic managers and schedulers in determining when to schedule booked patients. These general rules of thumb are listed below.

1. Schedule booked patients at the start of the clinic day. This allows for booked patients to be served while the backlog of walk-in demand has yet to increase significantly.
2. Schedule booked patients to correspond with time periods that typically have few walk-in patient arrivals.
3. Schedule booked patients during the periods of the day that the secondary provider is available to serve walk-in patients. This allows the primary provider to serve booked patients without the walk-in patients being neglected.
4. Booked patients should not be scheduled in the time immediately following a lunch break. This allows for the backlog of walk-in patients that have been accumulating during the lunch break to be served.
5. The last booked patient should be scheduled such that a sufficient amount of time is remaining in the clinic day to serve the remaining backlog of patients.

An additional point to consider is how changes in scheduling methods will affect patient behaviour. Walk-in patient arrival patterns may begin to change as members of the community begin to recognize times throughout the day that they will be seen with a short waiting time. These adverse outcomes of changing scheduling methods chould be considered and walk-in arrival rates regularly reviewed.

7.1. Summary of Contributions

This research has studied the outpatient scheduling problem and has developed a method for evaluating schedule templates and finding a pseudo-optimal template for a given clinic setting. The contributions of this research are as follows:

- Developed a methodology allowing outpatient clinics to determine an appropriate volume of patients to book, and when to book them throughout the day
- Combined simulation and simulated annealing in a methodology that determines a clinic specific pseudo-optimal schedule template

- Developed a simulation model of the clinic using non-commercial software to evaluate the performance of a schedule template
- Developed a tool that can be used to evaluate results and allow users to adjust model parameters

7.2. Future Research

Future research has been identified to broaden the analysis of the outpatient scheduling problem at hand. A main benefit of the current model is the ability to develop results that account for the environmental factors unique to a given clinic. The model can easily be modified to account for unique walk-in arrival rates, appointment length distributions, and performance metric requirements. The logic used to model the flow of patients through the clinic is more difficult to adjust. This leaves the opportunity for future work that develops a framework for easily modifying the simulation conceptual model logic, such that the model is more adaptable to clinics with atypical patient flows. Additionally, this research has excluded patient no-shows and patient punctuality from analysis. These environmental factors can easily be added to the model in future work to broaden the analysis.

In this research, the effect that a given schedule template has on the daily performance of the clinic has been analyzed. This analysis could be taken a step further by observing the impact that schedule templates and booked patient volumes have on the access time of patients to the clinic. This would mainly be a function of the volume of booked patients the clinic is able to see in a day.

Chapter 8 References

[1] S.-H. Kim, "A Data-Driven Model of an Appointment-Generated Arrival Process at an Endocrinology Outpatient Clinic," [Online]. Available: http://www.columbia.edu/~ww2040/appointment_INFORMS_103115.pdf. [Accessed 15 October 2017].

[2] D. Gupta and B. Denton, "Appointment scheduling in health care: Challenges and opportunities," *IIE Transactions,* vol. 40, no. 9, pp. 800-819, 2008.

[3] J. Ross, "The Patient Journey Through Emergency Care in Nova Scotia," October 2010. [Online]. Available: https://novascotia.ca/dhw/publications/Dr-Ross-The-Patient-Journey-Through-Emergency-Care-in-Nova-Scotia.pdf. [Accessed 28 July 2017].

[4] L. McNamara, "Measuring Emergency Care Network Population Coverage with Location- Allocation Models and Geographic Information Systems," Dalhousie University, Halifax, 2017.

[5] S. Consulting, "CareRightNow Evaluating the Collaborative Emergency Centre Experience in Nova Scotia," 2014. [Online]. Available: https://novascotia.ca/dhw/publications/Care-Right-Now-Evaluating-the-CEC-Experience-in-Nova-Scotia-Full-Report.pdf. [Accessed 28 July 2017].

[6] N. T. 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),* vol. 14, no. 2, pp. 185-199, 1952.

[7] D. Lindley, "The Theory of Queues with a Single Server," *Mathmematical Proceedings of the Cambridge Philosophical Society,* vol. 48, no. 2, pp. 277-289, 1952.

[8] T. Cayirli and E. Veral, "Outpatient Scheduling in Healthcare: A Review of Literature," *Production and Operations Management,* vol. 12, no. 4, pp. 519-549, 2003.

[9] N. T. Bailey and J. D. Welch, "Appointment Systems in Hospital Outpatient Departments," *The Lancet,* vol. 1, no. 6718, pp. 1105-1108, 1952.

[10] A. Soriano, "Comparison of Two Scheduling Systems," *Operations Research,* vol. 14, no. 3, pp. 388-397, 1966.

[11] T. Cayirli and E. D. Gunes, "Outpatient appointment scheduling in presence of seasonal walk-ins," *Journal of the Operational Research Society,* vol. 65, no. 4, pp. 512-531, 2014.

[12] B. Fries and V. Marathe, "Determination of optimal variable-sized multiple-block appointment systems," *Oper Res,* vol. 29, no. 2, pp. 324-345, 1981.

[13] X. Liu and L. Liu, "Block appointment systems for outpatient clinics with multiple doctors," *Journal of the Operational Research Society,* vol. 49, no. 12, pp. 1254-1259, 1998.

[14] C.-J. Ho and H.-S. Lau, "Minimizing Total Cost in Scheduling Outpatient Appointments," *Management Science,* vol. 38, no. 12, pp. 1750-1764, 1992.

[15] C. Schaak and R. C. Larson, "An N-Server Cutoff Priority Queue," *Operations Research,* vol. 34, no. 2, pp. 257-266, 1986.

[16] E. J. Rising, R. Baron and B. Averill, "A Systems Analysis of a University-Health-Service Outpatient Clinic," *Operations Research,* vol. 21, no. 5, pp. 1030-1047, 1973.

[17] T. F. Cox, J. P. Birchall and H. Wong, "Optimising the queuing system for an ear, nose and throat outpatient clinic," *Journal of Applied Statistics,* vol. 12, no. 2, pp. 113-126, 1985.

[18] K. J. Klassen and R. Yoogalingam, "Appointment system design with interruptions and physician lateness," *IJOPM,* vol. 33, no. 4, pp. 394-414, 2013.

[19] T. Cayirli, E. Veral and H. Rosen, "Assessment of Patient Classification in Appointment System Design," *Production and Operations Management,* vol. 17, no. 3, pp. 338-353, 2008.

[20] J. Vissers and J. Wijngaard, "The outpatient appointment system: Design of a simulation study," *European Journal of Operational Research,* vol. 3, no. 6, pp. 459-463, 1979.

[21] M. J. Blanco White and M. C. Pike, "Appointment Systems in Out-Patients' Clinics and the Effect of Patients' Unpunctuality," *Medical Care,* vol. 2, no. 3, pp. 133-141, 1964.

[22] M. Samorani and L. R. LaGanga, "Outpatient appointment scheduling given individual day-dependent no-show predictions," *European Journal of Operational Research,* vol. 240, no. 1, pp. 245-257, 2015.

[23] K. J. Klassen and R. Yoogalingam, "An assessment of the interruption level of doctors in outpatient appointment scheduling," *Operations Management Research,* vol. 1, no. 2, pp. 95-102, 2008.

[24] K. J. Klassen and T. R. Rohleder, "Scheduling outpatient appointments in a dynamic environment," *Journal of Operations Management,* vol. 14, no. 2, pp. 83-101, 1996.

[25] K. J. Klassen and T. R. Rohleder, "Outpatient appointment scheduling with urgent clients in a dynamic, multi-period environment," *International Journal of Service Industry Management,* vol. 15, no. 2, pp. 167-186, 2004.

[26] S. Su and C.-L. Shih, "Managing a mixed-registration-type appointment system in outpatient clinics," *Int J Med Inform,* vol. 70, no. 1, pp. 31-40, 2003.

[27] X. Qu, Y. Peng, J. Shi and L. LaGanga, "An MDP model for walk-in patient admission management in primary care clinics," *International Journal of Production Economics,* vol. 168, pp. 303-320, 2015.

[28] M. Brahimi and D. J. Worthington, "Queueing Models for Out-Patient Appointment Systems -- A Case Study," *The Journal of the Operational Research Society,* vol. 42, no. 9, pp. 733-746, 1991.

[29] S. Kim and R. E. Giachetti, "A Stochastic Mathematical Appointment Overbooking Model for Healthcare Providers to Improve Profits," *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans ,* vol. 36, no. 6, pp. 1211-1219, 2006.

[30] S. W. Au-Yeung, U. Harder, E. J. McCoy and W. J. Knottenbelt, "Predicting patient arrivals to an accident and emergency department," *Emergency Medicine Journal,* vol. 26, no. 4, pp. 241-244, 2009.

[31] A. Kolker, Queuing Theory and Discrete Events Simulation for Health Care: From Basic Processes to Complex Systems with Interdependencies, Hershey: IGI-press Global, 2010.

[32] A. M. Law, Simulation Modeling and Analysis (Fourth Edition), New York: McGraw-Hill, 2007.

[33] N. Kortbeek, M. E. Zonderland, A. Braaksma, I. M. Vliegen, R. J. Boucherie, N. Litvak and E. W. Hans, "Designing cyclic appointment schedules for outpatient clinics," *Performance Evaluation,* vol. 80, pp. 5-26, 2014.

[34] M. d. S. G. Tsuzuki and T. d. C. Martins, Simulated Annealing Strategies, Potential Uses and Advantages, New York: Nova Publsihers, 2014.

[35] F. Busetti, "Simulated annealing overview," 2003. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=0F7C353EEFC4B639AD586BA2B017195F?doi=10.1.1.66.5018&rep=rep1&type=pdf. [Accessed 2017].

[36] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, "Optimzation by Simulated Annealing," *Science,* vol. 220, no. 4598, pp. 671-680, 1983.

[37] J. Banks, J. S. Carson, B. L. Nelson and D. M. Nicol, Discrete-Event System Simulation Fifth Edition, Upper Saddle River: Pearson Education Inc., 2010.

[38] S. Canada, "Census Profile," 2011. [Online]. Available: http://www12.statcan.gc.ca/census-recensement/2011/dp-pd/prof/details/page.cfm?Lang=E&Geo1=DPL&Code1=120052&Geo2=PR&Code2=01&Data=Count&SearchText=tatamagouche&SearchType=Begins&SearchPR=01&B1=All&Custom=&TABID=1. [Accessed 11 October 2017].

[39] S. Robinson, "Conceptual Modeling for Simulation," in *Proceedings of the 2013 Winter Simulation Conference*, 2013.

[40] M. J. Miller, N. Shahi and R. LaVecchia, "Allocating Outpatient Clinic Services Using Simulation and Linear Programming," in *Proceedings of the 2008 Winter Simulation Conference*, 2008.

[41] P. VanBerkel and J. T. Blake, "A comprehensive simulation for wait time reduction and capacity planning applied in general surgery," *Health Care Management Science,* vol. 10, no. 4, pp. 373-385, 2007.

[42] J. C. Lowery, "Getting Started in Simulation in Healthcare," in *Proceedings of the 1998 Winter Simulation Conference*.

APPENDIX A: Model Formulation

This research uses a combination of simulation and simulated annealing to develop pseudo-optimal schedule templates. This appendix will provide additional detail into the formulation of these modelling approaches. The notation used is shown below.

Notation

Symbol	Description
p	Patient (booked or walk-in)
d	Replication # (day)
D	Number of replications
α_d	Volume of booked patients that arrive to their appointment on day d
β_d	Volume of walk-in patients that arrive on day d
$A_{p,d}$	Arrival time of patient p
$S_{p,d}$	Start time of patient p
$L_{p,d}$	Appointment length of patient p
$E_{p,d}$	End time of patient p
θ	End of scheduled day for primary provider
S	Start time of the clinic day
F	End time of the clinic day

General Overview

The simulation and simulated annealing procedure work together to determine when to book patients (i.e. the pseudo optimal schedule template represented by $A_{p,d}, p \in [1,2, \dots, \alpha_d]$) such that the objective function:

$$\alpha * \sum_{d=1}^{D} \left(\frac{\sum_{p=1}^{\alpha_d+\beta_d} (S_{p,d} - A_{p,d})}{\alpha_d + \beta_d} \right) + \beta * \frac{\sum_{d=1}^{D} \max\left[0, \max\left(E_{\alpha_d,d}, E_{\alpha_d+\beta_d,d}\right) - \emptyset\right]}{D},$$

is minimized. Where $A_{p,d}$ is restricted to opening hours of the clinic (i.e. $S = 8\ am$, $F = 8pm$),

$$S \leq A_{p,d} \leq F \qquad \forall\ p \in [1,2, \dots, \alpha_d + \beta_d], \forall\ \ d \in [1,2, \dots, D]$$

And where the average waiting time per patient is less than or equal to 30 minutes,

$$\sum_{d=1}^{D}\left(\frac{\sum_{p=1}^{\alpha_d+\beta_d}\left(S_{p,d}-A_{p,d}\right)}{\alpha_d+\beta_d}\right) \leq 30\ minutes$$

And where the average over time is less than or equal to 10 minutes,

$$\frac{\sum_{d=1}^{D} \max\left[0, \max\left(E_{\alpha_d,d}, E_{\alpha_d+\beta_d,d}\right) - \emptyset\right]}{D} \leq 10\ minutes,$$

And where the schedule template $A_{p,d}$ is the same for each day of the simulation,

$A_{p,1} = A_{p,d} \quad \forall\ d\ \forall A_{p,d}$ values generated by the simulated annealing procedure.