A DATA-DRIVEN APPROACH FOR REDUCING PATIENT WAITING TIMES IN WALK-IN CLINICS

by

Vivekanand Anandhan, B.E.

A thesis submitted to the Graduate Council of
Texas State University in partial fulfillment
of the requirements for the degree of
Master of Science
with a major in Engineering
August 2017

Committee Members:

Eduardo Perez, Chair

Clara Novoa

Jesus Jimenez

ACKNOWLEDGEMENTS

I would first like to thank my thesis advisor Dr. Eduardo Perez in the Ingram School of Engineering at Texas State University. The door to Prof. Perez's office was always open whenever I had a question about my research or writing. He consistently allowed this paper to be my own work, but steered me in the right the direction whenever he thought I needed it. I would also like to thank the experts who were involved in the experimental phase of this research project: Dr. Clara Novoa, Associate Professor, Ph.D. and Dr. Jesus Jimenez, Associate Professor, Ph.D. Without their passionate participation and input, the research work could not have been successfully conducted.

I would also like to acknowledge Dr. Vishu Viswnathan in the Ingram School of Engineering at Texas State University as the second reader of this thesis. I am gratefully indebted to his very valuable comments.

Finally, I must express my very profound gratitude to my parents, fiancée, and to my friends and family for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this thesis. This accomplishment would not have been possible without them. Thank you.

Vivekanand Anandhan.

TABLE OF CONTENTS

LIST OF TABLES

Table **Page**

LIST OF FIGURES

Figure **Page**

ABSTRACT

Walk-in clinics have grown in popularity in the United States as a substitute for traditional medical care delivered in primary care clinics and emergency rooms. Walk-in clinics offer an affordable option for basic medical services when compared to a hospital emergency room or an urgent care clinic. This type of medical facility simplifies the health care process for many patients with non-life threatening conditions since no previous appointments are required to see a provider. However, the open access nature and lack of patient scheduling can lead to long wait times for patients or long periods of idle time for providers. In this thesis, we derive a *discrete-event* simulation model to study pure walk-in clinics where patients are served without appointments. In addition, a non-linear programming model is developed to capture the trade-offs between the clinic and patients benefits and costs. A case study is discussed that consider a walk-in clinic located in central Texas. The computational study provides useful insights that are applicable to any walk-in health care facility. For instance, a trade-off between management cost and patient satisfaction can be achieved by proper allocation of resources at each station of the walk-in clinic. Even with various levels of demand (low, normal, and high), it is possible for the clinic to achieve positive results. The analysis provides valuable guidance to clinic administration about allocation of resources to improve patient satisfaction and the overall clinic performance.

1. INTRODUCTION

The goal of this research is to develop models for improving patient service and resource management in walk-in clinic environments. This research proposes a methodology towards achieving this objective, including a simulation model for decision making purposes and an analytical model for evaluating the decisions.

This research is motivated by the increasing popularity of walk-in clinics in the US. A walk-in clinic is a medical facility that accepts patients on a walk-in basis and with no appointment required. Walk-in clinics provide basic medical services such as vaccinations, evaluation of flu symptoms, and treatment of some physical injuries. Several health care service providers fall under the walk-in umbrella including urgent care centers, retail clinics and even many free clinics or community health clinics. Walk-in clinics offer the advantages of being accessible but, since no appointments are provided, patients often experience long waiting times. Excessive patient waiting time has been identified as a primary source of overall patient appointment dissatisfaction among the general medical patient population. For most patients, time is precious and waiting time for treatment is a waste. Therefore, patients who wait for their treatment appointment feel more stress, confusion, and frustration, which affect their response to treatment [1]. Often, waiting periods are caused for reasons beyond patients' responsibilities. For example, sometimes patient documents are not processed in time for treatment or the provider capacity is not enough to service the patient demand. In such cases, patients may decide to stay and wait or may leave to seek service in another clinic.

Although patients may visit a walk-in clinic to seek treatment for different conditions, they are required to follow the same process to see a provider. The process can be divided into three steps. The first step of the process is the check-in at the front desk. At this point, patients submit their personal identification and insurance cards to the front desk staff. Patients can be categorized in two groups: new patients and existing patients. New patients need to complete extensive paperwork about their medical history before getting admitted to the clinic. The documents are completed in the waiting room and usually take the patients about twenty minutes. After the paperwork is done, new patients provide a co-pay and wait in the waiting area until they are called inside the clinic by the medical assistant. Existing patients are patients who have visited the clinic before and their information is already in the system. The paperwork they should complete to get admitted is limited and most of the time they are just required to provide a co-pay and wait for their turn to be called by the medical assistant.

The second step of the process is performed by one of the clinic medical assistants. In this step patients are transferred to a small room of the clinic where the patient's blood pressure, weight, and body temperature are measured. The medical assistant also gathers information about patient symptoms and condition. Once the assessment process is completed, the patient is transferred to the examination room. Then, the medical assistant reports the case history to the physician.

The third and final step of the process is the patient's medical examination. Before seeing the patient and entering the examination room, the physician/provider reviews the patient's case history. Once in the examination room, the time spent with each patient differs per the severity of the case. After visiting with the patient, the provider will go back to the office to enter the case information in the system and will come back to the patient examination room with a prescription if needed. Some patients might be required to undergo a medical test (blood or x-ray) before being discharged. The provider will notify the patient not requiring medical tests when it is time to check-out. The flowchart in Figure 1 demonstrates the overall operation of the health care system considered.

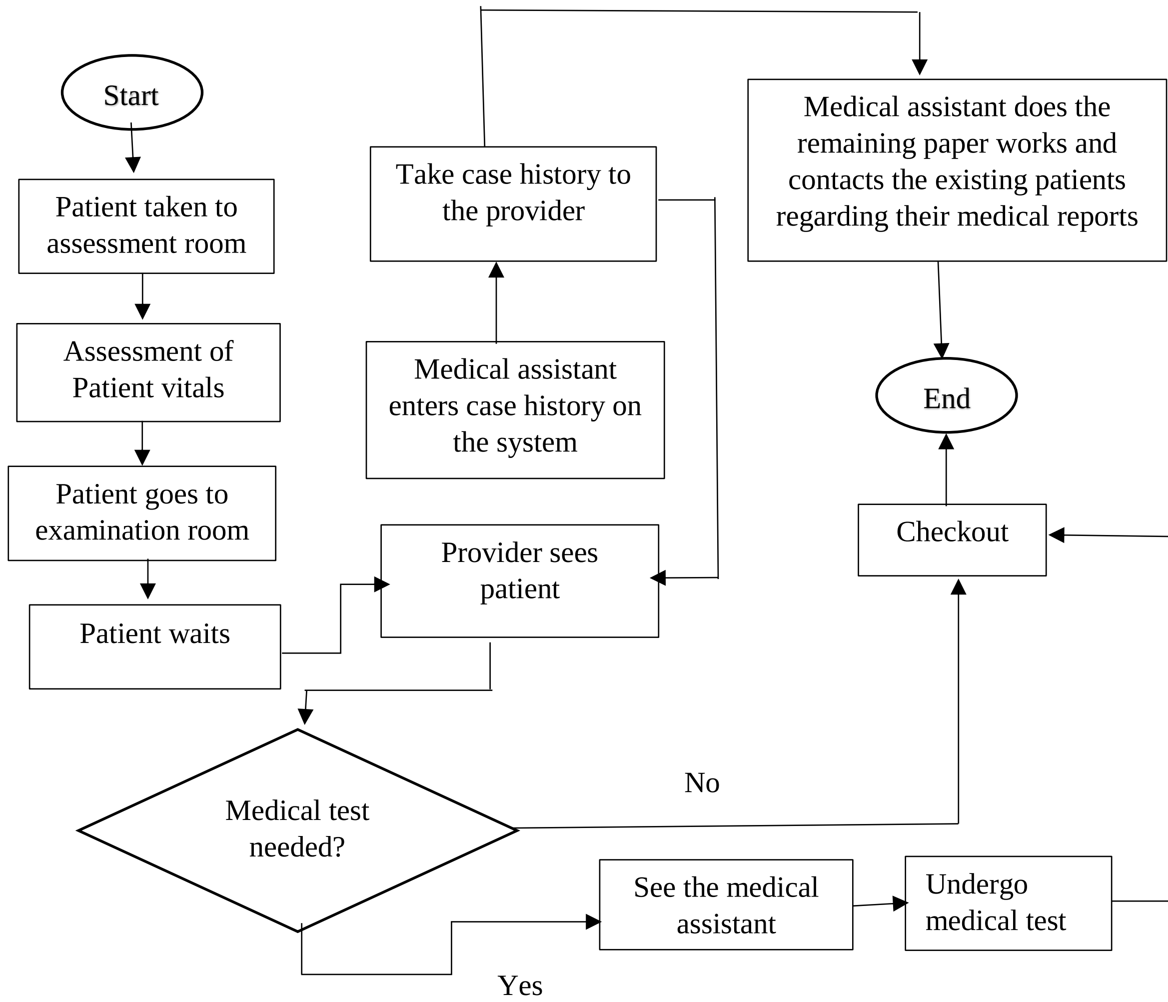

Figure 1: Overall Process Flowchart

Patient service management in walk-in clinics (requiring no patient appointments) is a very challenging problem that has received very little research attention. Current research has focused on developing scheduling techniques for patient appointments in outpatient clinics but no research has been published that considers the walk-in clinic environment. The remainder of the thesis is organized as follows. First, past literature related to this research is reviewed. The review demonstrates that operations in a walk-in clinic have not been extensively studied. In the third chapter, the thesis discusses about the simulation model followed by explanation of the non-linear programming model in the fourth chapter. Finally, in the fifth chapter, conclusion and directions for future research are offered.

2. LITERATURE REVIEW

Prior research on patient service management in walk-in clinics is very limited. Both simulation and optimization have been considered as viable approaches to patient management in outpatient clinics, open access clinics, and emergency rooms. These settings have some similarities with walk-in clinics. However, outpatient and open access clinics differ significantly from walk-in clinics because they admit patients with appointments. Emergency rooms differ considerably from walk-in clinics because of their availability of resources. The literature review discusses papers published in each of the three areas mentioned and illustrate the difference between them and walk-in clinics.

2.1. Outpatient Clinics

An outpatient clinic is where the patients are admitted based on scheduled appointments. The main issues considered when studying outpatient clinics is the patient waiting time, providers idle time and over time, and the overall cost of the procedure. There are many research works carried out for improving outpatient services including the scheduling of the patients to reduce waiting times and the proper management of the resources. Multiple methods have been used to optimize the scheduling in the outpatient clinic including Markov decision process (MDP) models, queuing theory, heuristic approaches, and discrete-event simulation.

Current methods for scheduling patients in outpatient clinics mostly rely on the use of the mean and standard deviation of service times, and no-show and walk-in probabilities. Haung *et al.* [2] proposed an alternative method by designing a predetermined scheduling template that accounts for patient wait time and physician idle time without overbooking and double booking in outpatient clinics. The method made sure that both patient wait time and the provider's idle time meet the declared scheduling policies without over booking. The authors concluded that the system can effectively reduce the patient waiting time as much as 56% without significantly increasing the provider's idle time. In addition, the new schedules will have same number of patients per clinic session and this ultimately enhances the quality of care. On the other hand, Zhu *et al.* [3] developed an efficient approach for outpatient scheduling by specifying a bidding method and converting it to a group role assignment problem. The approach was validated using simulation experiments. The work states that an efficient outpatient scheduling approach was obtained because of considering the outpatient scheduling as a collaborative activity and the creation of a qualification matrix for the application of the group role assignment algorithm. Later, Tugba *et al.* [4] developed an alternative appointment system through simulation using the universal dome rule and compared the new approach against some of the best traditional appointment rules in the literature. The results of the evaluation showed that the universal dome rule performed better in terms of decreasing the overall cost of the waiting time of the patient, and the idle and overtime of the doctor. A heuristic rule is proposed for identifying the optimal patient appointment sequence with accuracy.

Tang *et al.* [5] developed an appointment scheduling algorithm that considers routine and urgent patients taking into account the patient no–show probability. The goal of the algorithm was to reduce the weighted sum of average waiting time, physician waiting and over time. A heuristic algorithm based on two kinds of shifting policies (HE-TKS) is developed to solve the appointment schedule. Computational experiments were conducted to show how the critical factors affecting the service efficiency of the clinic. Similarly, Liu *et al.* [6] proposed heuristic dynamic polices for scheduling patient appointments. In this work, cancellation of appointments and patient no-show were considered to derive the policies. The results of this work reveal that the heuristics proposed outperformed all other benchmark policies, particularly during high patient load compared to the normal capacity. The authors suggest that the open access (OA) policy could also be a reasonable choice during low patient load.

In terms of resource utilization, Tsai and Teng [7] developed a stochastic overbooking model to enhance the service quality as well as to increase the utilization of multiple resources while considering the patient's call-in sequence. Two methods were proposed for the estimation of the expected waiting and overtime cost: the convolution estimation method and joint cumulative estimation method for computing expected waiting time and overtime cost with multiple resources. Similarly, Kim and Giachetti [8] used a stochastic mathematical over booking model (SMOM) that considered the probability of no-shows and walk-ins. The results of the paper stated that the SMOM can significantly increase the expected total profit under high and stochastic no-show rate. The sensitivity analyses demonstrated that the SMOM is robust under a diversity of health care environments and cost structures.

Patrick [9] developed a Markov Decision Model (MDM) for optimal outpatient appointment scheduling. His work demonstrated that a small booking window does significantly better when compared to open access services. The author demonstrated that over a wide variety of potential scenarios and clinics, the MDM policy did well when compared to OA in terms of minimizing costs as well as providing more consistent throughput. A similar work was carried out by Feldman *et al.* [10] to schedule appointment under patient preference and no show behavior. The objective was to maximize the expected net profit per day by developing a static model that does not consider the current state and by developing a dynamic model that considers the current state of scheduled appointments coupled with a heuristic solution approach. The computational experiments performed in this work revealed that the proposed policies perform well under a variety of conditions such patient no-shows and cancellations.

Zeng *et al.* [11] studied the problem of clinical scheduling with overbooking of heterogeneous patients. A local search algorithm is proposed to find the optimal schedules. LaGanga [12] considered the problem of capacity expansion in outpatient clinics. Her work demonstrated that no-shows can be managed effectively through consumer engagement along with creative use of overbooking while allowing the use of flexible capacity. Recommendations for effective and appropriate use of overbooking were provided. Mocarzel *et al.* described a simulation model for an outpatient healthcare clinic facing multiple issues related to patient admission and resource workflow. Factors affecting these services were identified and design of experiments was conducted and analyzed to evaluate how these factors and the factor interactions impact average waiting time at check-in, average number of patients waiting in queue, average waiting time.[13] Sowle *et al.* developed a framework for patient admission planning and intermediate term allocation of the resources. The framework was developed by integrating discrete-event simulation and Integer programming. The results of this work provided a trade-off between maximum number of patients that scheduled in a day versus their waiting time at the front desk. [14]. A similar work by Walker *et al.* developed a simulation based methodology for planning the schedules of providers and the appointment for patients. The results of this paper show that there is a trade-off between the maximum number of new patients that can be scheduled per appointment time at the clinic versus the patient waiting time at the front desk.[15].

Tang *et al.* [16] proposed an optimal schedule that takes no-show behavior into account to determine the time intervals between patients for minimizing the overall cost of the patient waiting time, physicians idle and overtime. Similarly, Ho and Lau[17] considered various appointment scheduling rules and investigate their ability to minimize a weighted sum of medical personnel's and patients' idle time costs. The idle times incurred by probability of no-show, coefficient of variation of service times, and number of patients per clinical session were taken as the three factors to run the experiments. These experiments were validated using simulation. Cayirili and Gunes [18] investigated the appointment system in outpatient clinic systems, as a combination of access rules and appointment scheduling rules, designed explicitly for dealing with walk-in seasonality. This study integrated capacity and appointment decisions which are usually addressed in an isolated manner. Simulation optimization was used to derive heuristic solutions to the appointment scheduling problem. Similarly, Guo *et al.* [19] came up with a triage strategy to guarantee patient care for critically ill patients while utilizing resources efficiently.

The literature on patient service management in outpatient clinics reveal that most of the research focuses on appointment scheduling algorithms and capacity allocation for which various methods and approaches have been handled such as the discrete-event simulation, MDP, heuristic rules and Lean Six Sigma. However, the process at the walk-in clinics are different as there is no prior appointments made and the flow of patients is uncertain.

2.2. Open access clinic

An open-access clinic admits both patient with appointments as well as walk-ins. The goal of open-access clinics is to utilize the provider's time efficiently and to provide service to patients without the need of appointments and with a minimum waiting time. There are multiple papers proposing ideas to improve the operations of open access clinics. Most of the papers rely on methodologies such as stochastic models, discrete-event simulation, and heuristic rules to provide interventions for the operations of these clinics under multiple scenarios.

Kopach *et al.* [20] stated that open access can lead to significant improvements in clinic throughput with little sacrifice in the continuity of care. This work focus on the effect of clinical characteristics on successful open access scheduling. The study was made with the help of a discrete-event simulation built using data drawn from a local clinic. The results showed that double booking had a significant effect in increasing the continuity of care in open access clinics while appointment lead time had the largest effect on the patient throughput. Robinson and Chen [21, 22] compared the traditional and open access appointment scheduling policies for a single provider. In 2010, the same group of authors, compared the performances of appointment policies for both traditional and open access clinics if the number of patients arriving to the clinics was different as per their distributions for no-shows or same day appointments. A hybrid scheduling policy was developed where open access policies dominated the traditional policies.

Peng *et al.* [23] studied the open access clinical scheduling problem with overbooking for heterogeneous patients. The authors developed a hybrid simulation and a genetic algorithm approach to determine the heuristic optimal scheduling templates for admitting walk-in patients. The results showed that, unlike the overbooking model for homogenous patients, the model for heterogeneous patients is not multi-modular i.e. the heuristic optimal scheduling template were significantly affected by patient attendance rate, level of demand for same day appointments and walk-ins, as well as the cost coefficients. Similarly, Qu *et al.* [24] proposed a hybrid open access policy adopting two time horizons for open scheduling and the research paper investigated a scenario where more than one time horizon for open appointments is justified. The analytical results showed that the optimized hybrid open access (OA) was better than the optimized current single time horizon OA policy in terms of expectation and variance of number of patients consulted.

La Ganga and Lawrence [25] proposed a research framework for appointment overbooking to improve patient service and clinic performance. In this paper, a flexible appointment scheduling model is constructed to mitigate the detrimental effects. A queuing analytical model of appointment scheduling was developed and a heuristic solution methodology was employed to find the best solution for wide range of problem settings. However, a general conclusion on how an overbooking schedule should be constructed is not drawn.

Health care quality may improve with short notice appointment schedules and with higher patient show-up rates. However, patient flow uncertainty and variability negatively impact the service design for open access clinics. La Ganga and Lawrence [26] examined the problem of no-shows and proposed appointment overbooking as one means of reducing the negative impact of no-shows. A new clinic utility function was developed to capture the trade-offs between the benefits and costs. The results from a series of simulation experiments reveal that the overbooking provided a greater utility when clinics served large number of patients, no-show rates were higher and, service variability was lower. Qu *et al.* [27] presented a closed form approach to determine the optimal percentage of open access appointments to match daily provider capacity to demand. The results of this paper demonstrated that the optimal percentage of open access appointments mainly depended on the ratio of the average demand for OA appointments to provider capacity and the ratio of the show-up rates for prescheduled and OA appointments. Later, in 2012 again Qu et al [28] demonstrated how to select the percentage of short notice or open appointments in an OA scheduling system. A mean-variance model was developed and an efficient solution procedure was derived to determine the open appointment percentage by increasing the average number of patients examined while reducing the variability. The numerical results from this work indicated that for cases with high patient demand and no-show rates for fixed appointments, one or more Pareto optimal percentages of open appointments significantly decreased the variability in the number of patients seen.

Lee and Yih [29] studied open access scheduling systems using simulation considering multiples scenarios. The scenarios included different levels in the demand variability, no-show rate, and ratio of same day patients. The results provided an insight about how to configure open access policies under different conditions. The work also elaborates on the effect of slot composition for the same day and on the effects of pre-booking on different clinical environments. The limitation was that the physician's service time was fixed. A study by Rose *et al.* [30] investigated the impact of advanced access scheduling on no-show rates, practice finances, patient satisfaction, and health care utilization. The results from the study revealed that for practices with high no-show rates, advanced access appeared to yield marked improvements. On the other hand, the research study concluded that the advanced access is less effective for practices with lower baseline no-show rates.

Qu and Shi [31] studied the impact of patient choice on the performance of provider capacity in open access clinics. A Markov's decision process model was developed for sequential clinical scheduling with the objective of improving patients scheduling to optimize clinic performance. The proposed approach estimated the performance of a practical capacity policy in a significantly short time. Lin et al [32] investigated the effect of appointment schedules on the accessibility and efficiency of open access clinic operations. A Markov Decision Processes (MDP) model was developed for sequential clinical scheduling. The developed model was solved by Dynamic programming (DP) for small problems. Approximate Dynamic Programming (ADP) algorithms based on aggression and simulation were developed to find schedules for larger problems. Chakraborty *et al.* [33] developed a sequential clinical scheduling method for patients with general service time distributions. The results of this work stated that the policy developed yields an objective evolution that provides a convenient stopping criteria and experimentally illustrate the effect of service time variance on clinic capacity.

Since they have similarities with walk-in clinics, the methodologies proposed for open access clinic management can provide insight on how to approach the walk-in clinic problem. However, walk-in clinics do not serve patient with appointments which makes their operation highly uncertain and challenging.

2.3. Emergency Room

The emergency room and the ambulatory care are the other important areas of research that are like walk-in clinics. The scheduling of the ER's is of prime importance when it comes to patient satisfaction. Lean six sigma processes along with simulation techniques are used to optimize the emergency room availability to keep the patient satisfied. Similarly, stochastic models are also developed to improve the operating room planning.

Although discrete-event simulation has made significant progress in the health care field, it has been tested primarily in hospitals and specialty clinics. Morrison and Bird [34] developed a model that considered the front office operation and patient care processes in ambulatory health care. Visual mapping and simulation tools were mentioned as some of the critical components that lead to successful outcomes like improved quality of care and overall process improvement. Similarly, Duguay and Chetouane [35] studied emergency department systems using discrete-event simulation. Their objective was to reduce the patient waiting time and to improve overall service delivery and system throughput. As waiting time is associated to the available resource, several alternatives were designed based on adding resource scenarios.

The challenges and opportunities for scheduling appointments in emergency department are discussed by Gupta and Denton [36]. The research work discusses relevant factors for operating emergency department including patient waiting time, information delivery, and the quality of service provided to the patient. Thompson *et al.* [37] discussed the effects of waiting times and perceived waiting times in emergency departments. The study concluded that projecting expressive quality and managing waiting time perceptions and expectations might be more effective strategies than decreasing the actual waiting time. Mowen *et al.* [38] reported a research that investigated patient perceptions of service, performance, quality, and patient satisfaction. This work concluded by stating that quality dimensions such as waiting time, availability and utilization of resources have an important role in determining the patient satisfaction.

Spaite *et al.* [39] a presented a rapid process redesign in an emergency department to reduce the waiting time intervals. The application of a process improvement team approach was used to evaluate and redesign patient flow. The authors showed that process redesign was possible in a complex, tertiary-care ED. These changes resulted in decrease in the waiting time intervals and improvements in patient satisfaction. The work by Mandahawi *et al.* [40] was also on reducing the waiting time at an emergency department and here design of six sigma has been implemented to develop a triage process for an emergency department. Discrete-event simulation models were developed and the results revealed that the length of stay and waiting time is reduced after the triage system was implemented. Jones and Schimanski [41] provided a systematic review of clinical outcomes with a four hour target to reduce emergency department waiting time. Ferrand *et al.* [42] provided a review and classification of the state of research on efficient utilization of operating rooms (ORs).

Lamiri *et al.* [43] derived a stochastic model for operating room planning with two types of demand for surgery, namely, elective surgery and emergency surgery. Elective surgery could be planned. Emergency is random in terms of arrival of patients and the problem was in assigning the elective cases to different periods over a planning horizon to minimize the sum of elective patient related costs and overtime costs of operating rooms. A stochastic mathematical programming was proposed in combination with Monte Carlo. The solution of this method was proved to converge to a real optimum as computational budget increases.

Ayvaz and Huh[44] focused on allocation of limited resources by formulating a general model and adopting a dynamic- programming approach. The results of this research work show that the system involving lost sales and backorders is not simple but exhibit desirable monotonic properties. Later, in 2015, Nunez *et al.* [45] focused on particular combination of human resources for each surgery. This research paper investigated surgery scheduling problems considering simultaneously, the operating room and post anesthesia recovery. An integer linear programming model that allowed to find optimal solutions for small size instances was proposed, and a meta-heuristic based on a genetic algorithm was developed that solved larger size instances.

3. SIMULATION ANALYSIS

The aim of this chapter is to develop a modelling approach that will help in reducing patient waiting times by identifying and implementing a balanced combination of resources in walk-in clinics. The clinical environment and the simulation model are discussed in detail in this chapter. A discrete-event simulation model was constructed using a real-time data collected from a walk-in clinic in Central Texas. The implementation and validation of the model was done in collaboration with the clinical and technical staffs from the walk-in clinic. The Rockwell Software ARENA simulation package was used to implement the simulation model. The results are discussed in detail and insights are provided for operation of walk-in clinics.

3.1. Clinic environments and system dynamics

As stated earlier, this research is motivated by the growing popularity of walk-in health care clinics in the United States. During this research, the team spent three months at Live-Oak walk-in care, a pure walk-in clinic located in Central Texas recording time-stamped data, observing the patient flow from arrival to check-out and discharge, and exploring the type of interactions that each staff member has with the patients. Thus, the team could construct process maps of the common paths through the clinic from multiple perspectives (patient and management). In addition, the team could derive probability distributions for most of the activities of the patient flow paths. This lead to the development of a discrete-event simulation model for walk-in clinics. The model was validated from the observational data and subsequently used to design interventions and policies to improve the system operations.

The working hours of the clinic are weekdays (Monday – Friday) from 9:00 AM to 11:59 AM for morning session and 12:00 PM to 7:00 PM for afternoon sessions. On weekends (Saturday and Sunday) the clinic operates from 10:00AM to 2:00 PM. The clinic had one front desk staff, two medical assistants and one provider. The clinic has one waiting room for patients, one assessment room that could accommodate only one patient at a time for whom the basic assessments were done, four examination rooms, and one X-ray room.

Patients arriving at the clinic are classified into two categories, (i) new patients and (ii) existing patients. The difference between these two categories was that new patients should complete a large sum of paper work before seeing the provider. This increased the length of stay of the new patients compared to the existing patients. Figure 2 shows a bull's eye chart that focuses on various aspects of the clinic's performance. The chart was prepared based on the results of a survey obtained from the patients and their feedback over two quarters of a year. The percentile was calculated by keeping the performance of other clinics under the management as a benchmark.

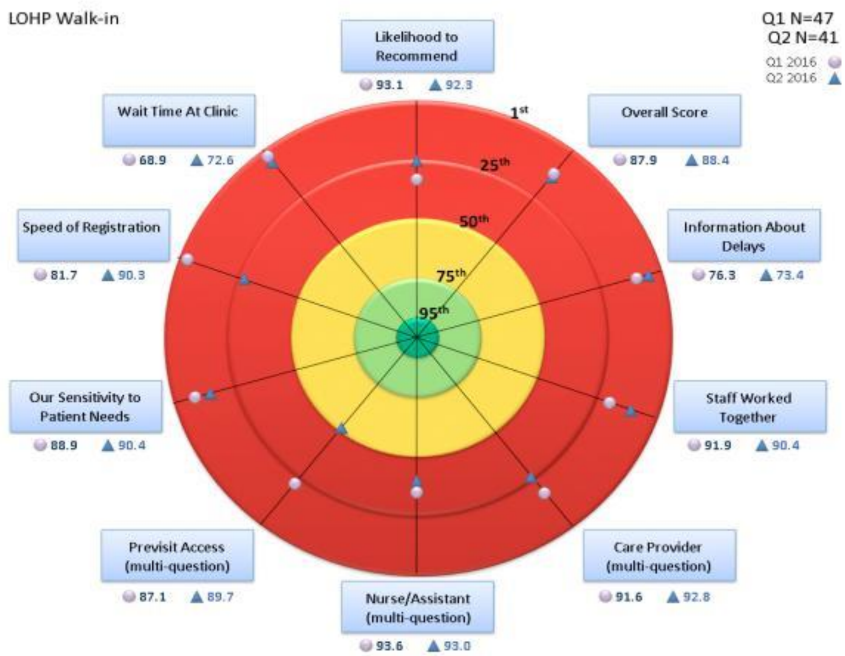

Figure 2: Bull's eye chart of the walk-in care

The bull's eye chart provides a means of visualizing the simultaneous progress toward each goal. The chart is divided in percentiles. The highest the percentile value (green area), the better the performance of the clinic. The circles and triangles represent the clinic's performance for the first and second quarter of 2016 respectively. The Live-Oak has a lot of room for improvement since the values for the performance measures are within the lower percentiles for both initial quarters of 2016. In terms of patient waiting time, the clinic is currently performing within the 1st percentile.

There are three main factors that served as input for the simulation model. Firstly, the patient inter-arrival times (patient demand) per session on weekdays and weekends. Similarly, the medical assistants' service times at the clinic and finally the provider's service times per session from Monday to Friday and on weekends. Table 1 shows the probability distributions like gamma (GAMM) for patient arrival in the morning, normal

(NORM) for medical assistant service time in the afternoon, log-normal (LOGN) in weekends and so on.

Table 1: Distribution of input factors

Input Factor	Morning	Afternoon	Weekends
Patient inter arrival times (minutes)	GAMM (11.3, 1.14)	GAMM (10.5, 1.23)	LOGN (18.2, 21.7)
Medical assistant service times (minutes)	GAMM (1.66, 2.46)	NORM (5.62, 2.02)	LOGN (2.75, 2)
Provider service time (minutes)	NORM (29.3, 20.7)	BETA (0.628, 1.1)	TRIA (-0.001, 10.8, 98)

3.2. Simulation model

The simulation model section describes each segment of the process in the clinic from patient check-in to patient check-out. First, the arrival of the patients was simulated using the create block where the distribution of patient arrival is entered. In the next step, as discussed in the previous section the patients are identified as new and existing patients and it is simulated by a decision block in which the percentage of the new and existing patients are entered. Secondly, the waiting time of the patient is simulated using a hold block. The hold block holds the entities (patients) until the given condition is satisfied. In this model, the condition was to release an entity when the assessment room is empty (equal to 0) and if one of the examination room is available (exam room < 4). Thirdly, the process at the medical assessment room and the examination room is simulated by using a process block. Finally, simulation of the process at the X-ray room is performed by sorting out the patients into two types namely, (i) patients requiring medical test (ii) patients who do not need medical test. A decision block was used to simulate this decision process and the check-out (i.e., entities leaving the system) was simulated using the dispose block. Table 2 provides information about the percentage of new and existing patients, percentage of patients who needed medical test and those who did not require a medical test in the morning, afternoon and weekend scenarios.

Table 2: Percentage of patients in each session

	Morning	**Afternoon**	**Weekend**
New patients	78.12	42.86	43.47
Existing patients	21.87	57.14	56.52
Patients who needed medical test	28	42	60
Patients who did not need medical test	72	58	40

Figure 3 shows the overall design of the simulation model. Sub models 1, 2, and 3 represent the front desk, medical assessment, and examination room process respectively. Figure 4 shows the patients' arrivals which are generated by using a create block. Attributes and variables associated in the model are given in the assign block. The hold block resembles the waiting of the patient in the clinic. Figure 5 shows the process in the assessment room and at the examination room. Similarly, figure 6 shows the model built for the simulation of the medical test process at the X-ray room.

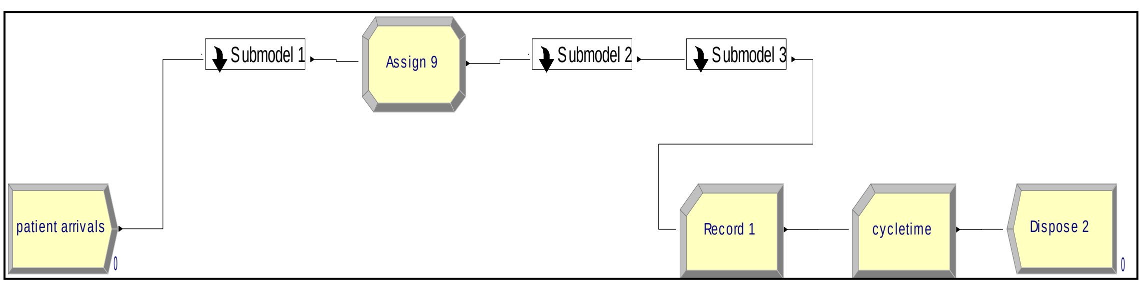

Figure 3: Overall simulation model of walk in clinic

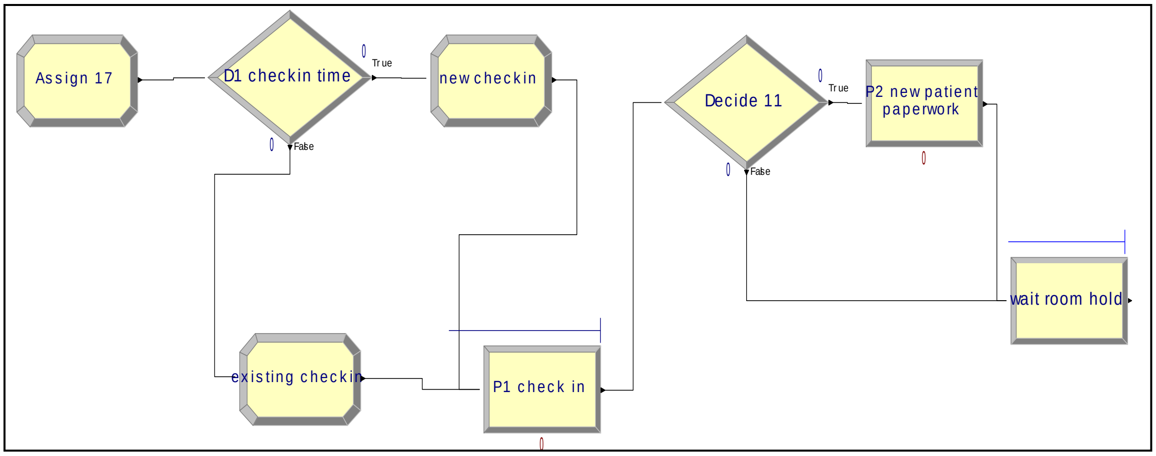

Figure 4: Front desk simulation model

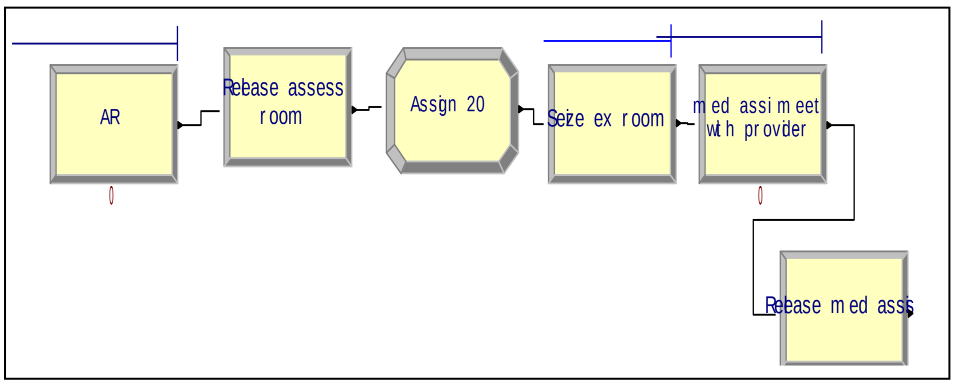

Figure 5: Medical assessment area simulation model

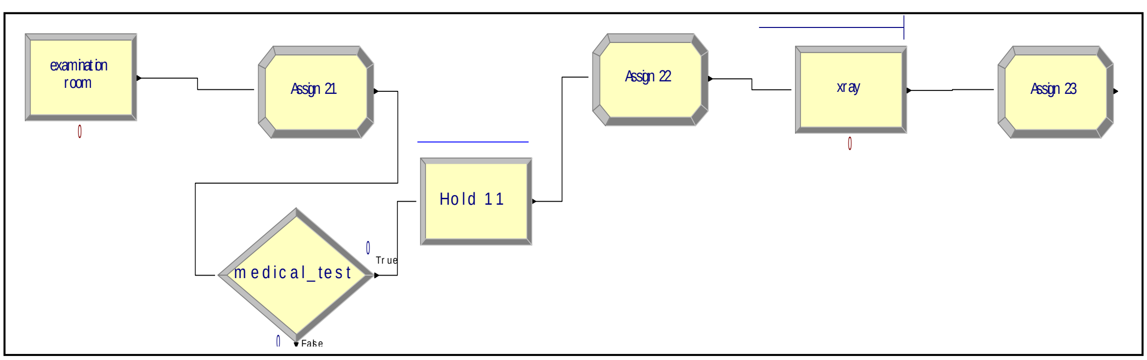

Figure 6: Examination room simulation model

3.3. Simulation validation

Theoretical calculations were used to compare the number of patients served per clinical session and the overall cycle time from the data obtained from the clinic. Then a simulation model was built for each scenario in such a way that it represents the exact clinical setup. The distribution of arrival rate to the system and process time at each station was obtained by using the SIMAN/ARENA input analyzer. The results obtained from the simulation runs were compared with the theoretical results to validate the model. Table 3 demonstrates the results from this validation step.

Table 3: Results from ARENA and Excel

	Morning		**Evening**	
	Arena	**Excel**	**Arena**	**Excel**
No. of patients served	17.7	15.3	17.4	19.56
Cycle Time	100.25	98.01	86.88	89.00

The model had 20 replications and each replication ran for 10 weeks. The results from the simulation model and the clinic are within a 10% difference. The results show that the simulation model represents the actual performance of the walk-in clinic and were presented to the clinic management and the simulation model was approved.

3.4. Design of experiments

A general factorial experiment was considered in this study. The results of the computational study provide a valuable insight on the positive and negative factors affecting the system. Figure 7 shows the inputs and outputs considered in the experiments.

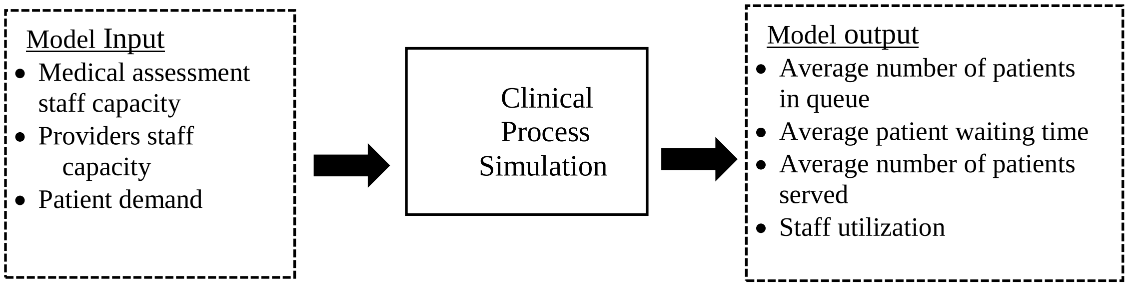

Figure 7: Input and output considered in the simulation study

The low, normal, and high levels for the factors (medical staff capacity, provider capacity, and patient demand) are given in Table 4, Table 5 and Table 6. Table 4 shows the levels and factors for the morning scenario which considers the clinic operation from 8:00 am to 11: 59 am. Table 5 demonstrates the levels and factors for the evening scenario which considers the clinic operation from 12:00 pm to 5:30 pm and finally, Table 6 reveals the levels and factors for the weekends which considers the clinic operations on Saturdays and Sundays from 9:00 am to 1:30 pm.

Table 4:Experimental factors and corresponding levels for the morning scenario

Factors	Low	Normal	High
Medical assessment staff capacity,	1	2	3
Providers staff capacity		1	2
Patient demand	"14%"	current distribution	"55%"

Table 5: Experimental factors and corresponding levels for the afternoon scenario

Factors	Low	Normal	High
Medical assessment staff capacity,	1	2	3
Providers staff capacity		1	2
Patient demand	"16%"	current distribution	"60%"

Table 6: Experimental factors and corresponding levels for the weekend scenario.

Factors	Low	Normal	High
Medical assessment staff capacity,	1	2	3
Providers staff capacity		1	2
Patient demand	"15%"	current distribution	"50%"

The list of experiments performed for the morning, afternoon and weekend scenarios is given in table 7

.Table 7: List of experiments

Experiment	Medical assessment staff capacity	Providers staff capacity	Patient demand
1	L	L	L
2	L	L	N
3	L	L	H
4	L	N	H
5	L	N	L
6	L	H	L
7	N	L	L
8	H	L	L
9	L	H	N
10	N	N	N
11	N	N	L
12	N	N	H
13	N	L	H
14	N	L	N
15	N	H	N
16	L	N	N
17	H	N	N
18	N	H	L
19	H	H	H
20	H	H	L
21	H	H	N
22	H	L	N
23	H	L	H
24	H	N	H
25	L	H	N
26	N	H	H
27	H	N	L

3.5. ANOVA

ANOVA was conducted to gather insights about the responses and to identify those factors that are significant for the experiments considered in the computational study. The factors observed in the experiments are:

- A = medical assistant capacity
- B = provider capacity
- C = patient demand

The responses (R) observed are as follows:

- Average number of patients in waiting room (R1)
- Average patient wait time in waiting room (R2)
- Average patient waiting time in exam room (R3)
- Average number of patients served (R4)
- Average utilization Med Assist 1 (R5)
- Average utilization Med Assist 2 (R6)
- Average utilization Med Assist 3 (R7)
- Average utilization Provider 1 (R8)
- Average utilization Provider 2 (R9)
- Average new patient cycle time (R10)
- Average existing patient cycle time (R11)

The following sections provide a discussion of the results of the analysis of experiments at a significance level that was assumed to be 5%. Sections 3.5.1, 3.5.2, and 3.5.3 discuss the results for the morning, afternoon and weekend scenarios respectively.

3.5.1. Computational results for the morning scenario

Figure 8 show the half-normal plot for response (R1) (average number of patients in the waiting room). Based on the plot, the significant factors for (R1) are the number of providers (B), the patient demand (C), and the interaction of those two factors (B and C). Response (R2), showed a similar behavior as the provider capacity and patient flow influenced waiting time of patients in the queue.

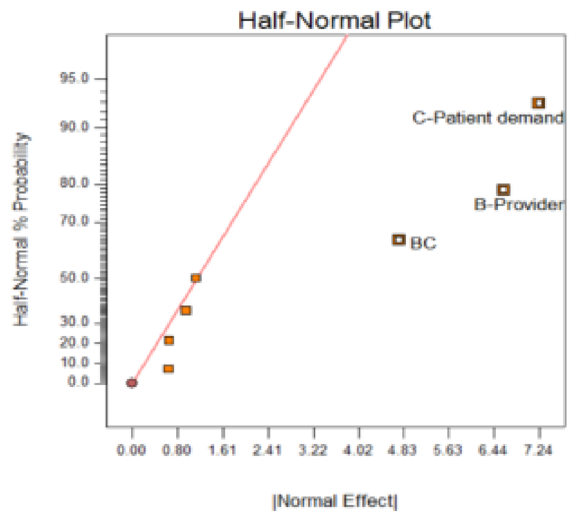

Figure 8: Average number of patients in the waiting room (R1) response half-normal plot

The responses R3 (average patient waiting time in examination room), average utilization of providers 1 and 2 (R8 and R9), average new patient cycle time (R10) and average existing patient cycle time (R11) showed a similar behavior. For all these responses, the factor that proved to be significant was the provider capacity (B). Figure 9 shows the half normal plot for response R3. Since, the other responses R8, R9, R10, and R11 also have provider capacity as significant factor, the half-normal plot is like that of response R3.

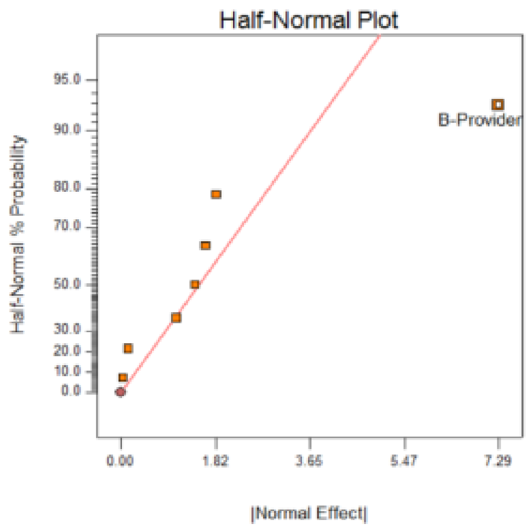

Figure 9: Average patient waiting time in the examination room (R3) response half-normal plot

The significant factors for response R4 (average number of patients served) are the provider capacity (B) and patient demand (C). Figure 10 depicts the half-normal plot for response R4.

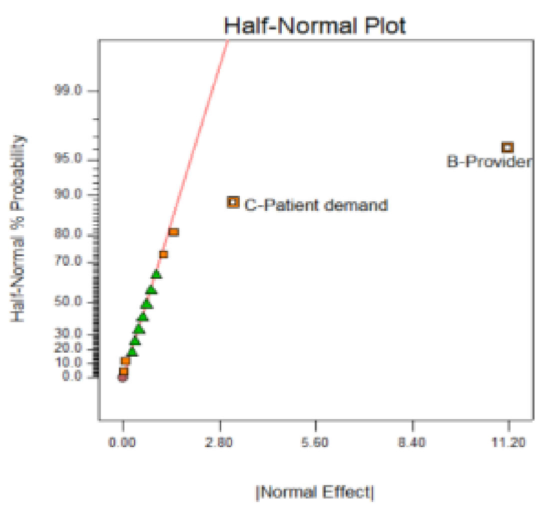

Figure 10: Average number of patients served (R4) response half-normal plot

Figure 11 shows the half-normal plot for response R5. For responses R5 and R6 (average utilization of medical assistants), medical assistant capacity (A), provider capacity (B), and the interaction of factors both (A and B) are the significant factors. The utilization of the provider and medical assistant are one of the significant factors for an improved clinical setting as it has an impact on the patient waiting time. The higher utilization of the provider will keep the medical assistants as they should move forward with assessment of the next patient in the queue. Both the responses R5 and R6 exhibit a similar half-normal plot.

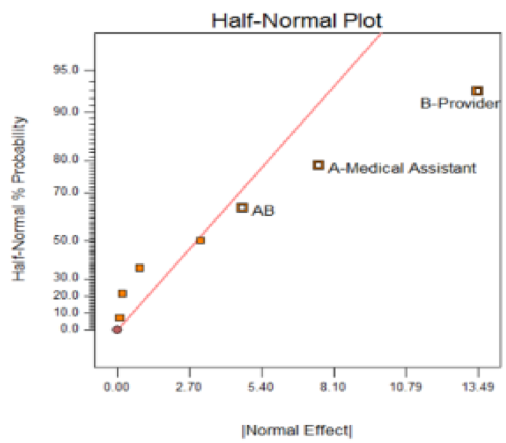

Figure 11: Average utilization of medical assistants (R5 and R6) response half-normal plot

Table 8 shows the standard deviation and mean for the responses obtained from the experiments for the morning scenario. The standard deviation and mean for each response was used to perform the ANOVA for the observations made in the morning scenario.

Table 8: mean and standard deviation for each response for morning scenario

Morning scenario			
		Mean	**S. D**
R1	Average number of patients in waiting room	2.15	1.22
R2	Average patient waiting time in waiting room	20.61	7.32
R3	Average patient waiting time in exam room	17.88	11.15
R4	Average number of patients served	7.41	2.35
R5	Average utilization of medical assistant 1	0.096	0.026
R6	Average utilization of medical assistant 2	0.04	0.021
R7	Average utilization of medical assistant 3	0.048	0.18
R8	Average utilization of provider 1	0.54	0.18
R9	Average utilization provider 2	0.27	0.24
R10	Average patient cycle time new	53.21	23.88
R11	Average patient cycle time existing	51.228	20.87

Figure 12 shows the different combinations of the factors taken into consideration and their corresponding results for the average patient waiting time in the waiting room and average number of patients waiting in the waiting room. In the morning session, it could be witnessed that there are nearly 3 or 4 patients waiting and their wait time was as high as 40 minutes before they are called by the medical assistants. The current combination (**NNN**), or benchmark, has one provider and two medical assistants and it impacts the waiting time of the patients in the waiting room. In-order to find the best combination for the morning session, a total of 27 experiments were conducted. The experiment 15 revealed that the best fit would be having a combination of two medical assistants (N) and two providers (H). This could reduce the number of patients waiting in the waiting room by two or less and their average waiting time reduced by 37%.

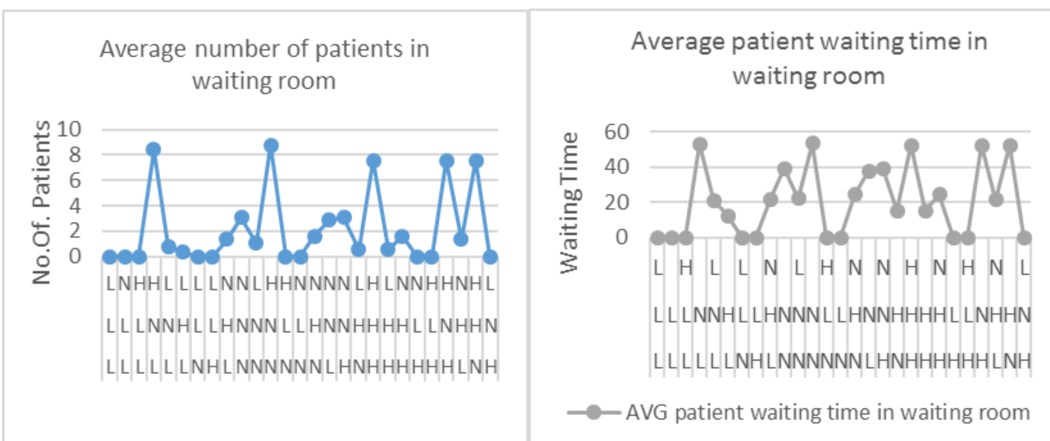

Figure 12: Average patient waiting and the number of patients waiting in the waiting room

The plot in figure 13 shows the average wait time of the patients in the examination room and the average number of patients served in that session. In the normal scenario where there is one provider, the waiting time at the examination room is as high as 55 minutes and the total number of patients served in the morning session seems to be 10 or less. The response from the experiments revealed that the average number of patients served has the provider capacity as one of its significant model term. From the set of experiments performed, by adding one provider (**NHN**), the number of patients served can be increased by 79% in the morning session.

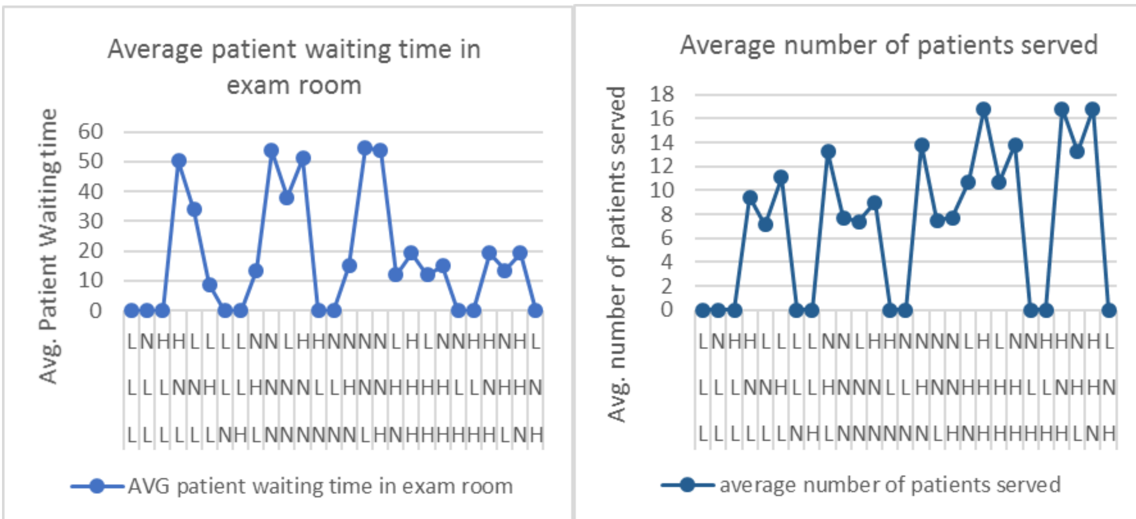

Figure 13: Average patients in the waiting room and the total number of patients served

The graphical distribution in figure 14 and figure 15 gives the result of average utilization of the medical assistants and the providers. The medical assistant capacity is set to one at low (L), two at normal (N), and three at high (H) levels. Similarly, the provider capacity is zero, one and two at low (L), normal (N) and high (H) levels, respectively. It is evident from the graph that for (**NHN**) the utilization of medical assistant 1 and 2 was 0.135 and 0.139 respectively and the provider 1 and provider 2 utilization was 0.84 and 0.79 respectively. This proves that the utilization of the resources during the **NHN** combination are almost equal during low, normal and high patient flow compared to the other experiments.

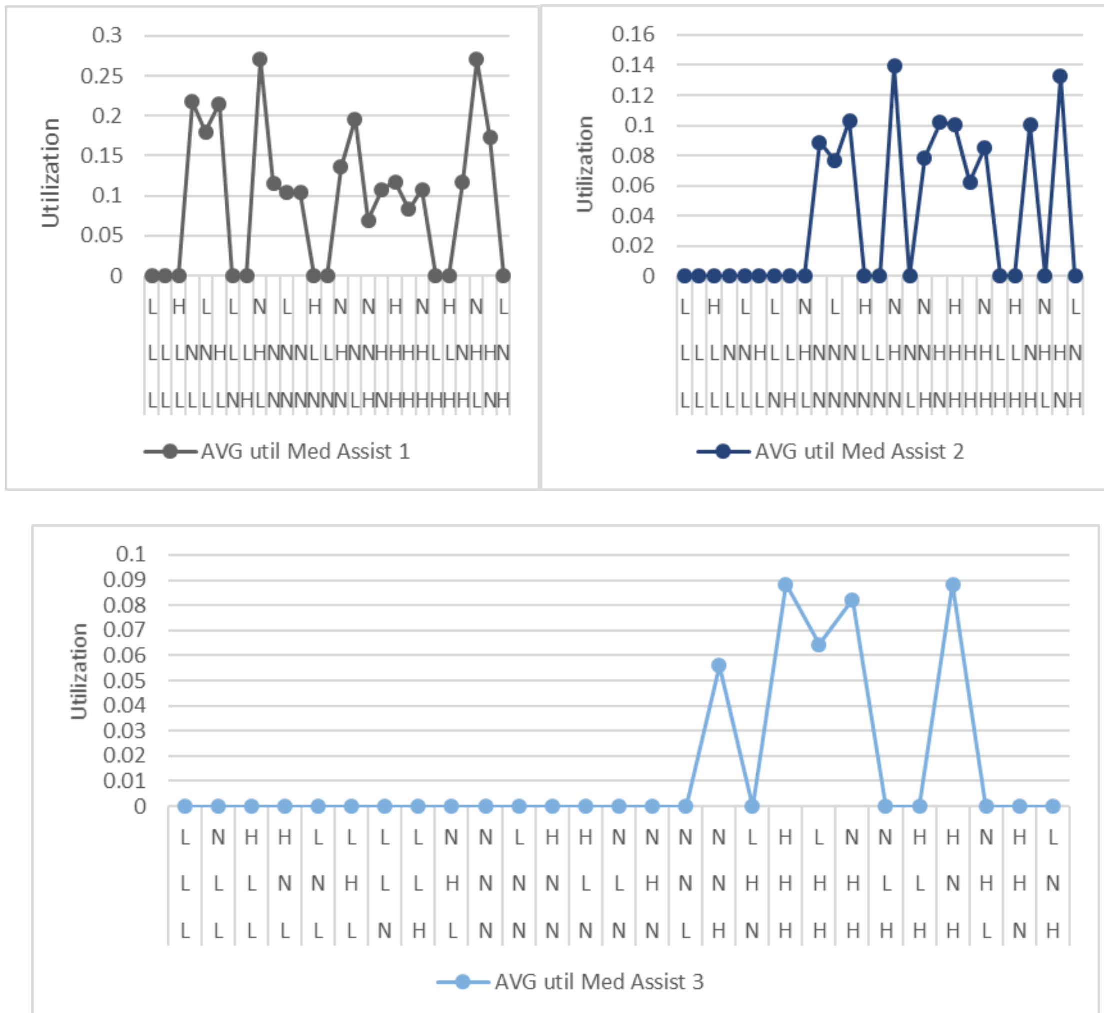

Figure 14: Average utilization of the medical assistants

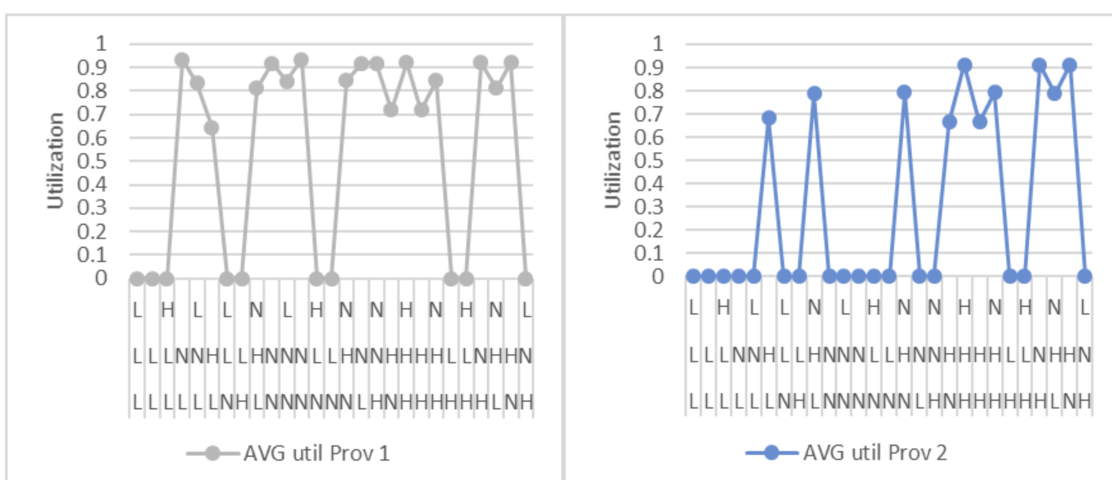

Figure 15: Average utilization of the providers

The cycle time for the new and existing patients vary because of the paperwork and it is also proportional with the provider and medical assistant capacity. Figure 16 shows cycle time for different combinations of the staff capacity, provider capacity and the patient demand. The optimum result is found when the provider is set to high (H) and the medical assistant is maintained at normal (N) (**NHN**). For this combination, the average cycle time reduces by 34% compared to the present cycle time in the clinic.

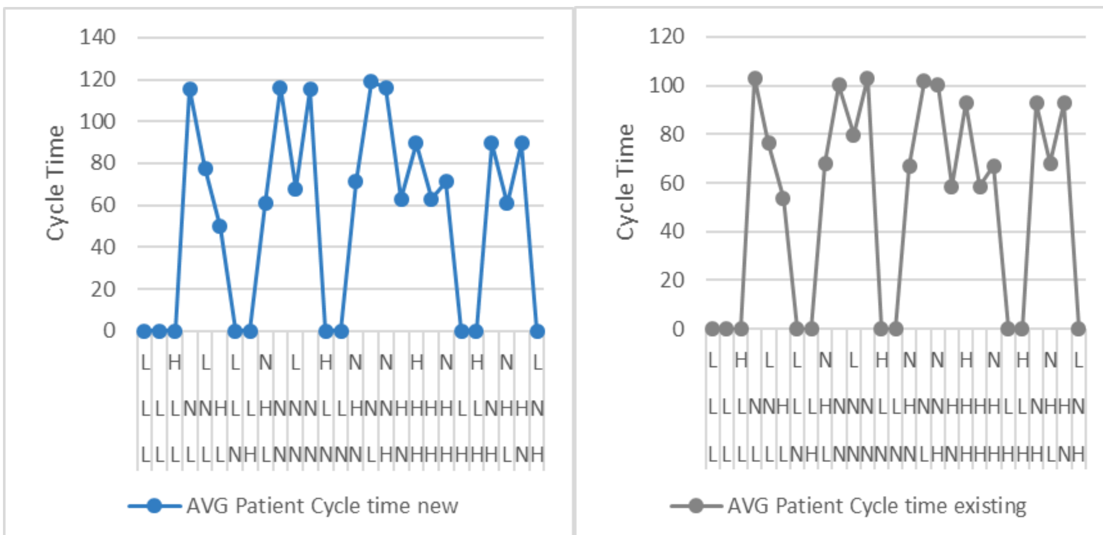

Figure 16: Average cycle time for new and existing patients

3.5.2.Computational results for the afternoon scenario

Responses (R1) average number of patients in the waiting room, (R2) average patient wait time in the waiting room, (R3) average patient waiting time in examination room, (R10) average new patient cycle time and (R11) average existing patient cycle time have the provider's capacity (B), patient demand (C), and interaction between the two factors (B and C) as the significant terms of the model. Figure 17 shows the half normal plot for response (R1). Responses (R2), (R3), (R10), and (R11) exhibit similar half normal plot as they are also dependent on the flow of patients and the capacity of the provider in the afternoon session.

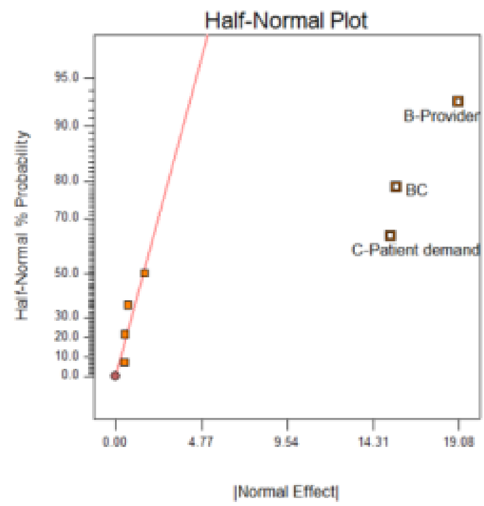

Figure 17: Average patient waiting time in waiting room (R1) response half-normal plot

Figure 18 and figure 19 shows the half-normal plot of response (R4) average number of patients served and (R6) average utilization of the medical assistant 2. These responses have provider capacity (B) as the significant model term. In the afternoon session, the second medical assistant was more available to communicate with the provider and the capacity of the provider regulated the patient flow and the utilization of the medical assistant 2.

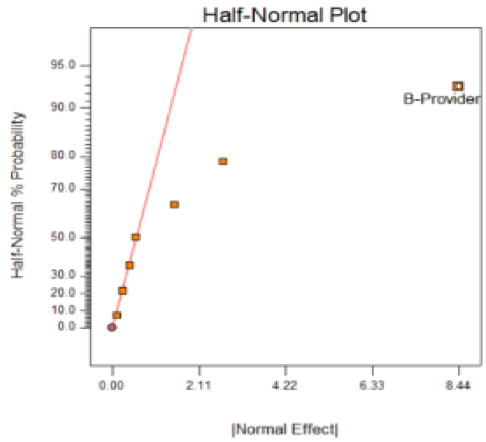

Figure 18: Average number of patients served (R4) response half-normal plot

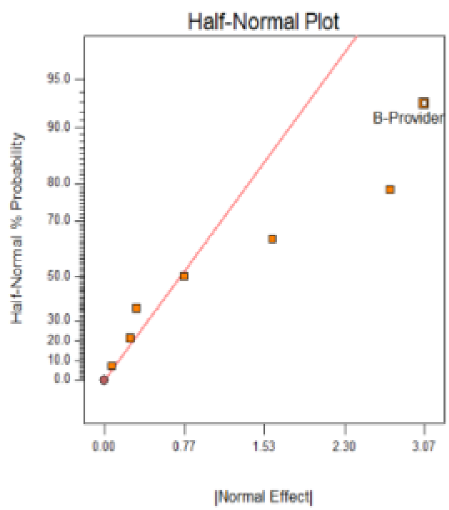

Figure 19: Average utilization of medical assistant 2 (R6) response half-normal plot

The average utilization of the medical assistant 1 (R5) depended on the provider capacity (B) and the medical assistant capacity (A) and the interactions of two factors (A and B) are the significant model terms for response R5. The half-normal plot for R5 is given in figure 20.

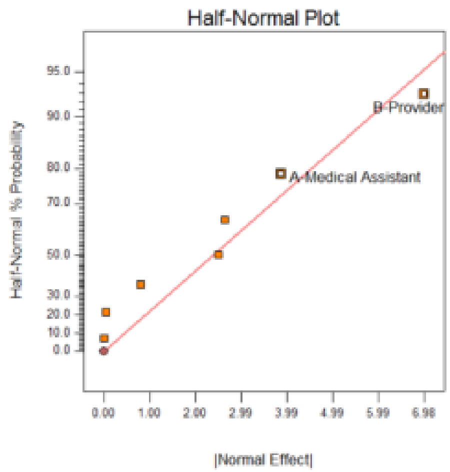

Figure 20: Average utilization of medical assistant 1(R5) response half-normal plot

The response (R8) average utilization of provider 1 has provider capacity (B) and patient demand (C) as the significant terms. The average utilization of the provider 2 (R9) has the provider capacity (B), patient demand (C) and their interaction (B and C) as the significant model terms. Figures 21 and 22 shows half normal plot for the response R8 and response R9, respectively.

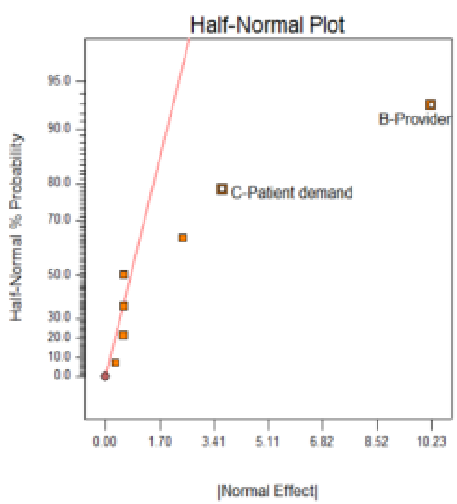

Figure 21: Average utilization of provider 1 (R8) response half-normal plot

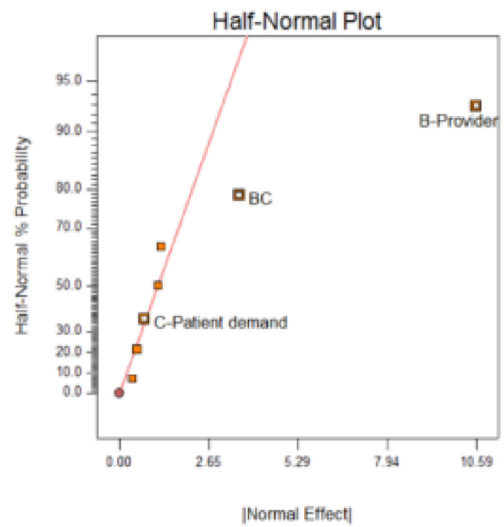

Figure 22: Average utilization of provider 2 (R9) response half-normal plot

Table 9 gives the detail about the standard deviation and mean for the responses from the set experiments run for the afternoon scenario. This standard deviation and mean of each response was used to perform the ANOVA for the observations made in the afternoon scenario.

Table 9: Mean and SD for each response in the afternoon session

Afternoon scenario			
		Mean	**S. D**
R1	Average number of patients in waiting room	1.38	0.43
R2	Average patient waiting time in waiting room	13.35	3.36
R3	Average patient waiting time in exam room	8.82	2.76
R4	Average number of patients served	12.73	5.65
R5	Average utilization of medical assistant 1	0.25	0.079
R6	Average utilization of medical assistant 2	0.076	0.08
R7	Average utilization of medical assistant 3	0.024	0.029
R8	Average utilization of provider 1	0.47	0.17
R9	Average utilization provider 2	0.23	0.13
R10	Average patient cycle time new	33.8	9.65
R11	Average patient cycle time existing	35.43	8.78

The graphical distribution of average number of patients in the waiting room and the time they wait in the waiting room is shown in the figure 23. The number of patients waiting in the waiting room is almost 2 and the wait time is more than 20 minutes with medical assistant capacity, provider capacity and patient demand to be at normal level (**NNN**). The number patients waiting in the waiting room could reduce to 80 % and the average waiting time can be reduced by 70% by adding a provider (**NHN**). This combination works well for the patient demand at any level (i.e., low, normal, and high).

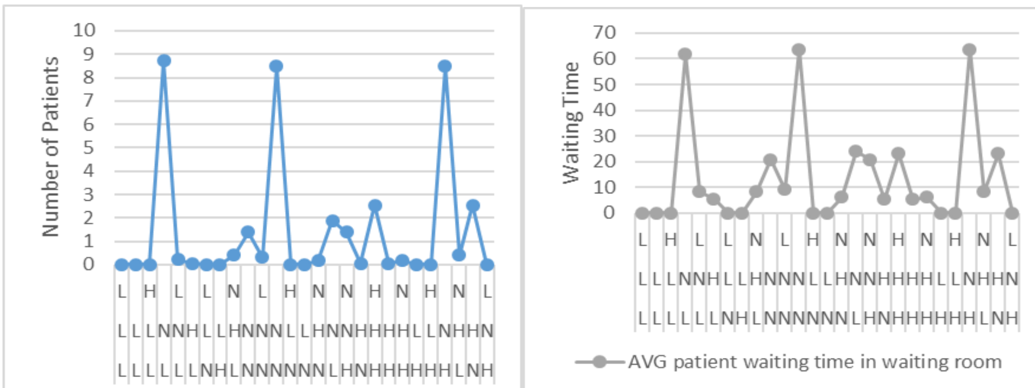

Figure 23: Average number of patients waiting and their waiting time in the waiting room

Figure 24 shows the plots of the average waiting time at the examination room, which is directly proportional to the provider capacity, and the average number of patients served in the afternoon. In experiment 12, where the patient demand is high and the capacity of the resources is normal or low (**NNH**), the waiting time is high and in turn the number of patients served reduces.

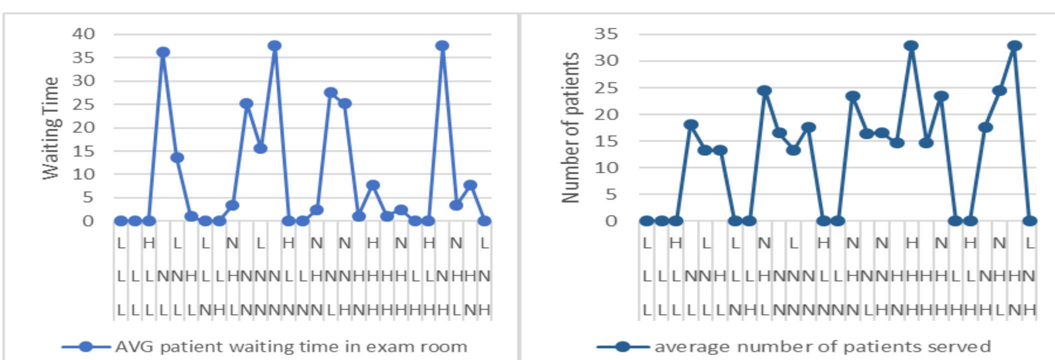

Figure 24: Average number of patients waiting in the exam room and average number of patients served

Figures 25 and figure 26 show the graphical representation of the average utilization of the medical assistants and the provider respectively for the afternoon session. From the plot for utilization of medical assistants it can be seen that the utilization is almost more or less equal for both the medical assistant 1 and 2 at normal level (N). The provider at normal (N) level has an utilization of 0.8 and when it is high (H) it can be seen that two providers are utilized equally with an utilization rate of 0.6 which is optimal as both the providers.

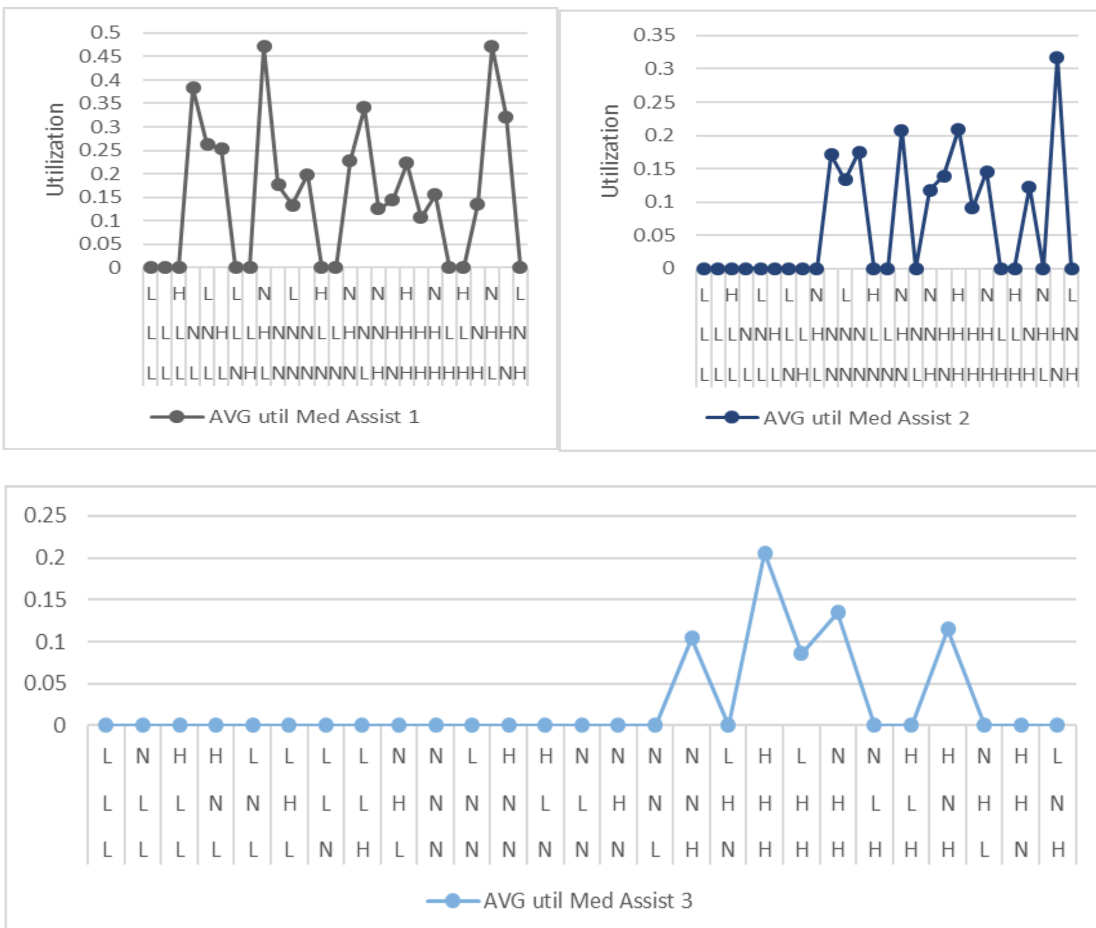

Figure 25: Average utilization of the medical assistants

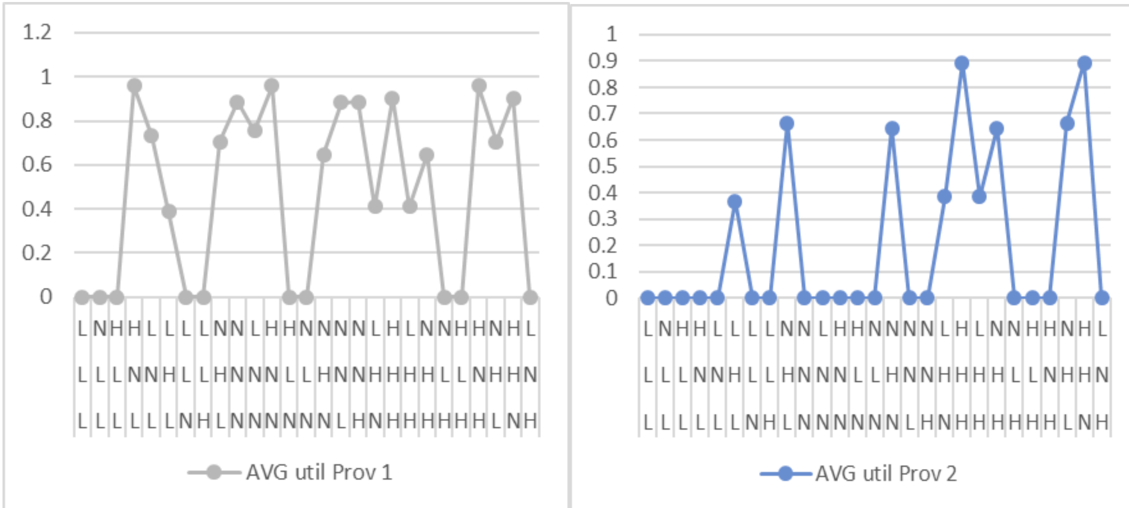

Figure 26: Average utilization of the providers

Figure 27 shows the graphical distribution of the average cycle time for the new patients and existing patients. With the current operating scenario of the clinic where there are two medical assistants and one provider, the cycle time observed is more than 60 minutes for both new and old patients. The optimum value of 50.33 minutes and 55.34 minutes for new and existing patients respectively are observed for the combination where the provider is maintained at high level whereas the medical assistants are maintained as normal level (**NHH**).

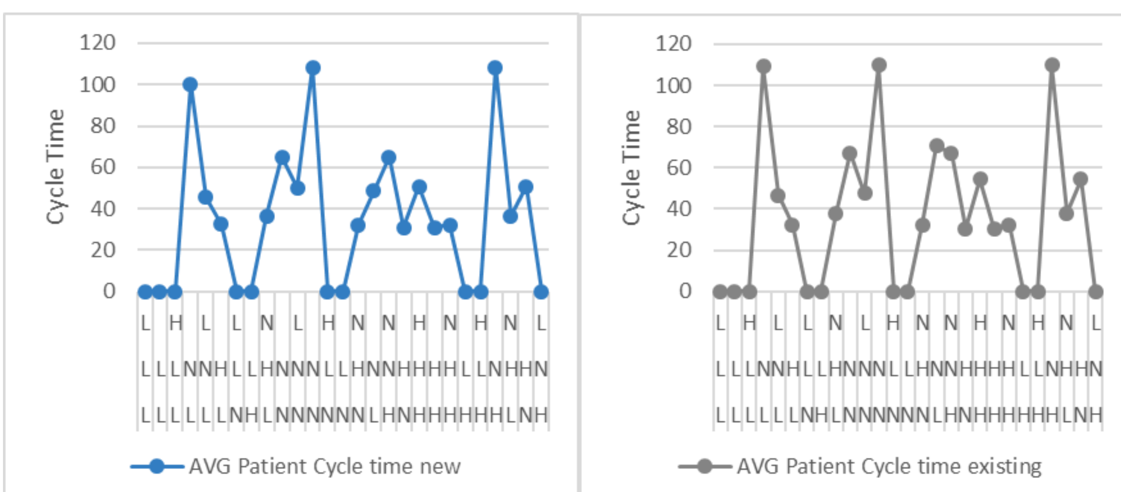

Figure 27: Average cycle time for new and existing patients

3.5.3. Computational results for the weekend scenario

Responses (R1) average number of patients in the waiting room, (R2) average patient waiting time in the waiting room, (R9) average utilization of provider 2 and (R10) average new patient cycle time have the factors provider's capacity (B), patient demand (C) and interaction between both the factors (B and C) as the significant model terms. Figure 28 shows the half normal plot for (R1). The other responses (R2), (R9) and (R10) have the same type of half normal plot as (R1).

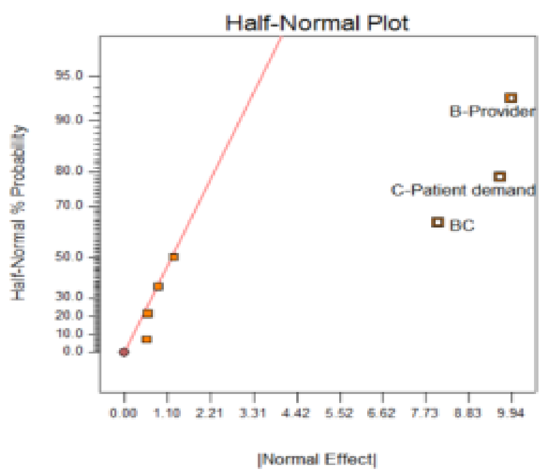

Figure 28: Average patient waiting time in waiting room (R1) response half-normal plot

The responses R3, (average patient waiting time in examination room), R4 (average number of patients served) and R11 (average existing patient cycle time) have the factors provider capacity (B) and patient demand (C) as significant model terms. Figure 29, shown below, is the half normal plot for the response (R4). The other responses (R3) and (R11) showed similar half-normal plots.

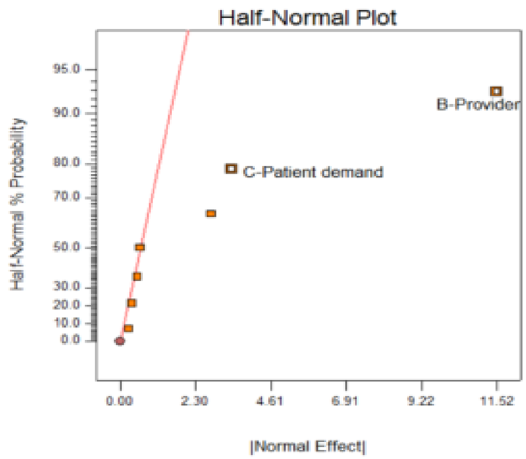

Figure 29: Average number of patients served (R4) response half-normal plot

The significant factors of response R8 (average provider 1) are the provider capacity (B) and its half normal plot is shown in the figure 32. The response R7, (average utilization of medical assistant 3) have medical assistant capacity (A), provider capacity (B) and their interaction (A and B) as the significant model terms and its half normal plot is shown in the figure 31. Similarly, the average utilization of medical assistant 1 (R5), has medical assistant capacity (A) and provider capacity (B) as their significant terms. Figure 30 shows the half normal plot of (R5).

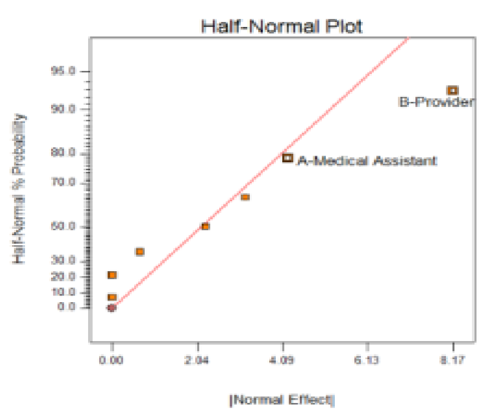

Figure 30: Average utilization of medical assistant 1 (R5) response half-normal plot

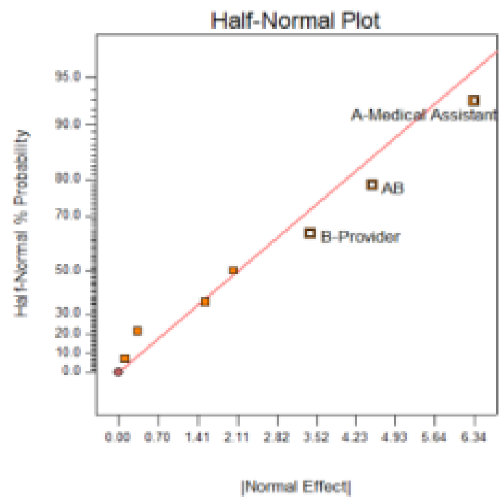

Figure 31: Average utilization of medical assistant 3 (R7) response half-normal plot

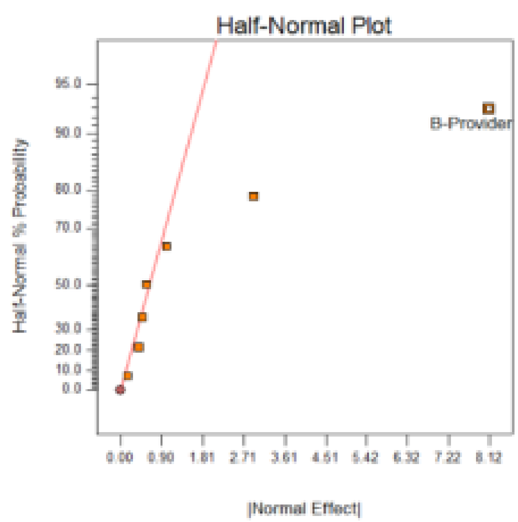

Figure 32: Average utilization of provider 1 (R8) response half-normal plot

Table 10 shows the standard deviation and the mean value for the eleven responses for the weekends scenario. This standard deviation and mean of each response was used to perform the ANOVA for the observations made on Saturday and Sunday (Weekends).

Table 10 : Mean and SD for the responses in the weekend scenario

Weekend scenario			
		Mean	**S.D**
R1	Average number of patients in waiting room	1.09	0.62
R2	Average patient waiting time in waiting room	11.01	4.80
R3	Average patient waiting time in exam room	12.73	7.33
R4	Average number of patients served	7.56	2.55
R5	Average utilization of medical assistant 1	0.06	0.02
R6	Average utilization of medical assistant 2	0.02	0.03
R7	Average utilization of medical assistant 3	0.09	0.01
R8	Average utilization of provider 1	0.46	0.21
R9	Average utilization provider 2	0.20	0.05
R10	Average patient cycle time new	35.71	12.54
R11	Average patient cycle time existing	34.31	15.65

The clinic experiences a relatively high patient flow on the weekends. The current system has one provider and two medical assistants to serve this high patient demand. It leads to high waiting time in the examination room that affects the overall cycle time of the patient. The distribution of the average number of patients waiting and their waiting time in the waiting room is given in figure 33. The combination (**NHN**) gives optimum value in terms of the waiting time and the number of patients waiting in the waiting room where the waiting time is reduced by 64% and the number of patients waiting is reduced by 90%.

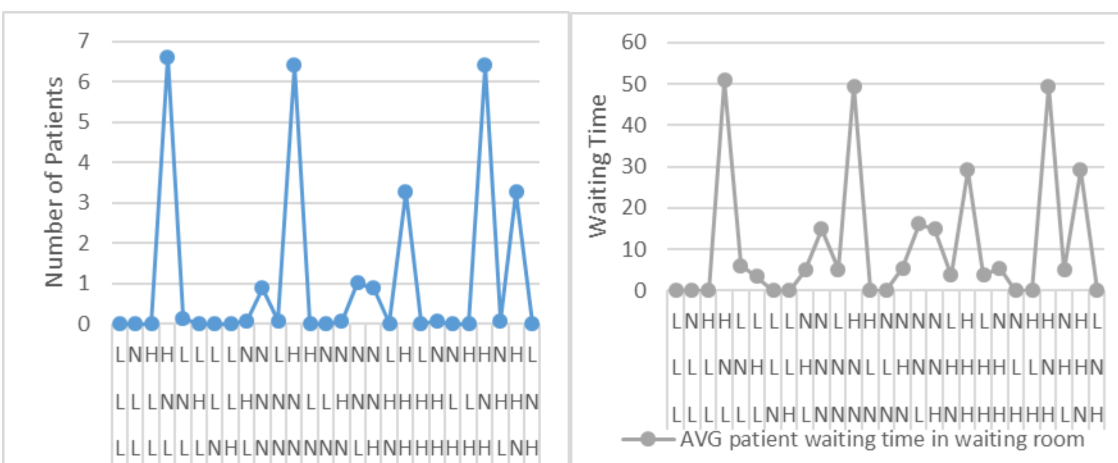

Figure 33: Average number of patients waiting and their waiting time in the waiting room

In figure 34, the results of the average patient wait in the examination room and average number of patient served is shown. It is evident that when the patient demand is normal (N) or high (H), by maintaining the provider capacity to be high (H) and medical assistant to be at normal (N) level, the average number of patients served per hour increases by 30%.

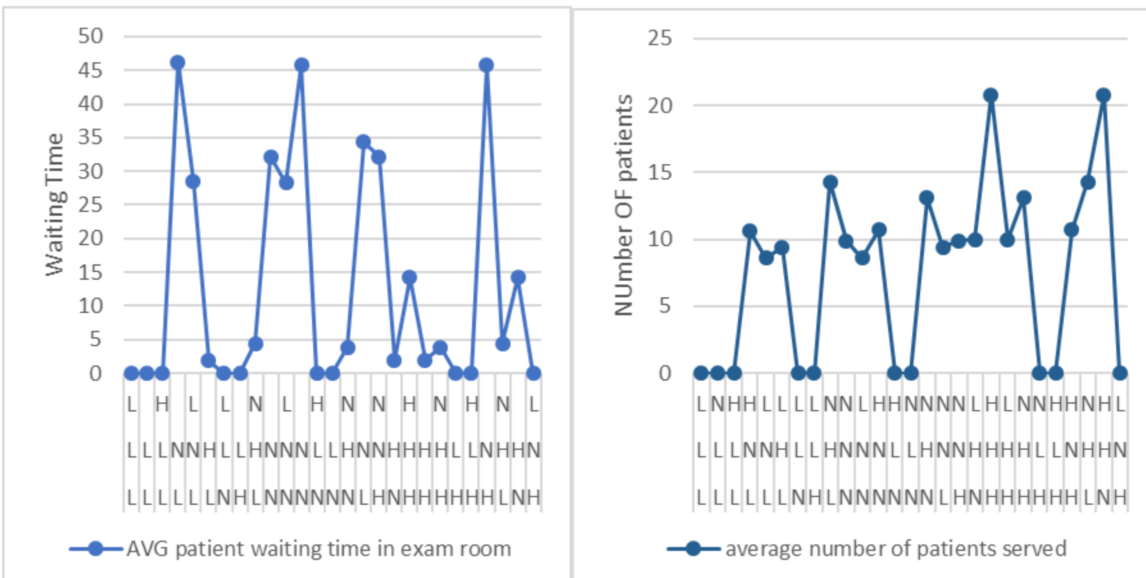

Figure 34: Average patient waiting time in the exam room and the average number of patients served

Figure 35 and 36 present the average utilization of the medical assistants and the providers respectively at low (L), normal (N) and high (H) levels. The utilization is based on the patient flow. The relatively high and equal utilization of the resources proves that the servers are allocated in an efficient manner. For instance, the utilization of providers 1 and 2 in experiment 15 is equal to 0.5, which proves that the providers are busy all day with reduced idle time compared to the other experiments.

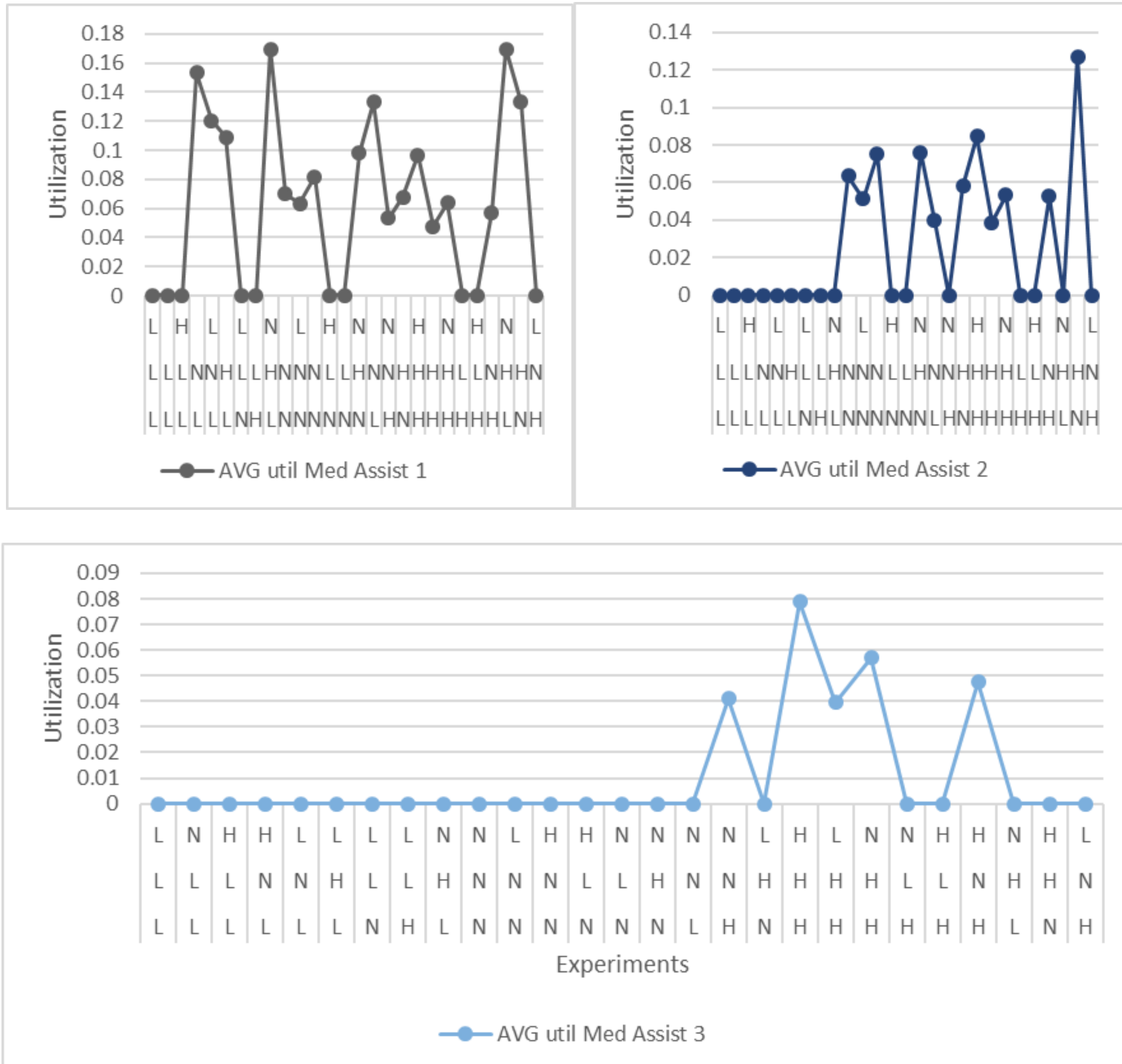

Figure 35: Average utilization of the medical assistants

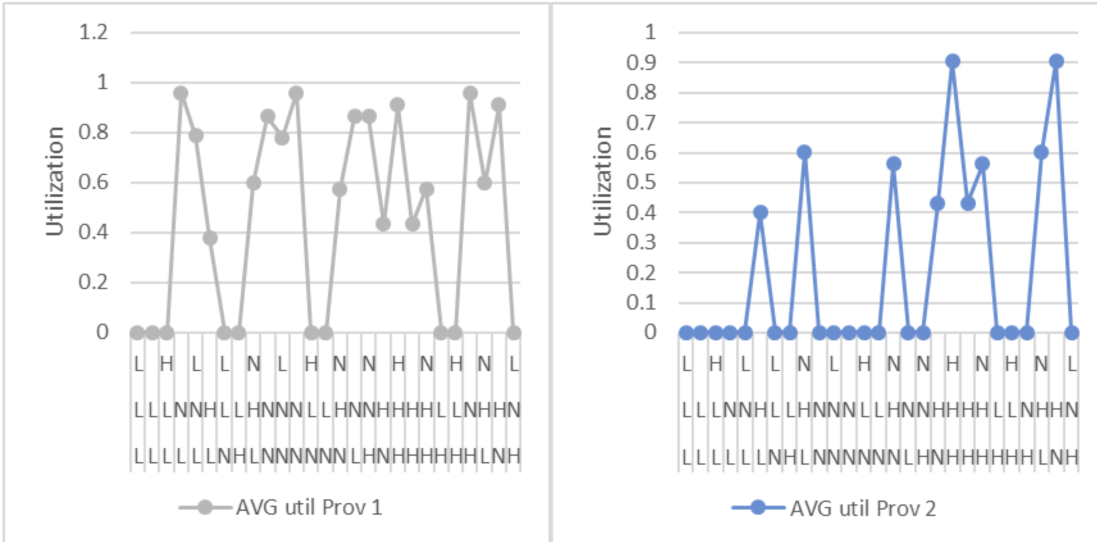

Figure 36: Average utilization of the providers

The plots for the average new and existing patient cycle time for the low (L), normal (N), and high (H) levels of the factors (medical assistants, providers, and patient demand) is given in the figure 37. If the medical assistant is kept at normal (N), provider at high (H), and patient demand is high (H) the values of the new and existing cycle time are 36.66 minutes and 30.22 minutes respectively which is 51% less than the present cycle time at the clinic. This is an optimal value as the combination (**N H N**) can serve a high patient flow with relatively less waiting time as compared to other combinations.

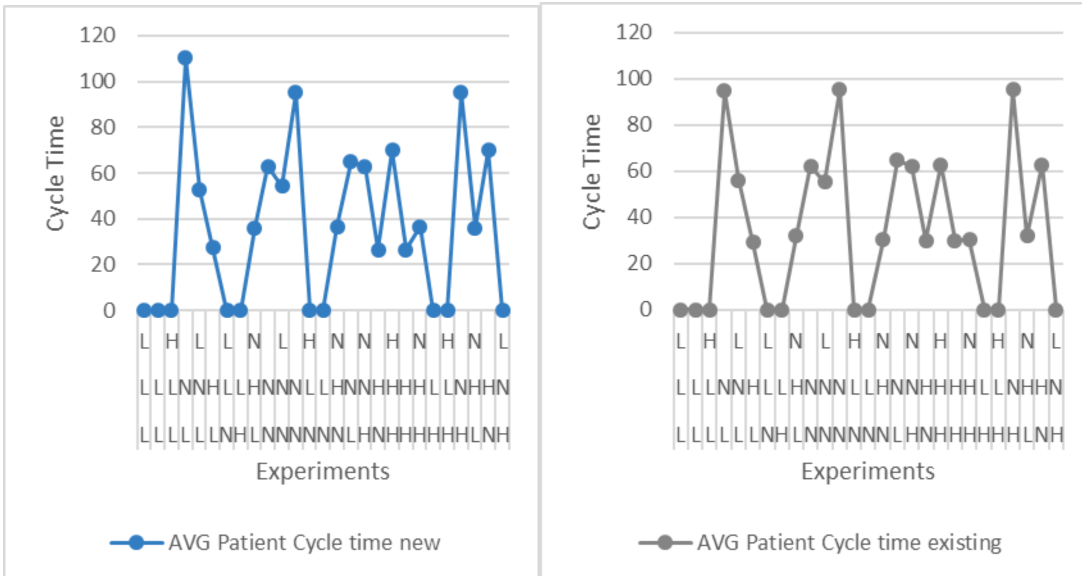

Figure 37: Average cycle time for new and existing patients

3.6. Discussion of the results

The walk-in clinics are one of the successful sectors in health care if operated in a more organized manner by allocating the resources efficiently. In this work, a walk-in clinic was observed and studied with discrete-event simulation model developed with SIMAN/ARENA simulation package. The computational results provide insights about how to manage resources at the clinic in a way it reduces the waiting time and length of stay of the patients.

For instance, the results show that, to reduce the waiting time in the waiting room at the clinic a second provider must be added to the clinic. A 38% reduction is observed in the waiting time between experiments 10 (NNN) and 15 (**NHN**) for the morning scenario. Similarly, a 70% and 64% improvement is observed in the afternoon and weekend scenario respectively. However, in the weekend scenario, the results (see experiment 12) also show that waiting time at the waiting room is higher than 40 minutes when the patient demand is increased by 15% (**H**) and one provider (**N**) serve at the clinic (figure 33). In this case, by adding two providers the waiting time will remain under thirty minutes and it is demonstrated in the results from experiment 26 (figure 33) in which the waiting time is reduced by 40% which is less than thirty minutes.

Regarding cycle time, from experiment 10 of the morning scenario which represent the actual system, the cycle time is more than an hour in the present system (**NNN**). This cycle time was reduced by 34% by increasing the provider capacity (experiment 15, **NHN**) to 2 (H). On the other hand, during high patient flow, as shown in experiment 26 **(NHH**), having two providers (H) reduced the cycle time by 11% compared to actual system. Upon studying this walk-in clinic, the above graphs reveal that adding one provider to the existing system would make the system efficient during high and normal patient demand. This results shows the benefit of having a simulation tool that will allow the clinic for planning for future expansion.

4. NON-LINEAR OPTIMIZATION MODEL

4.1 Problem and challenges

Walk-in clinics are primary care centers where the patients are admitted without a scheduled appointment. Walk-in clinics are a new trend in health care systems with the goal of increasing patient access to health services. The first walk-in clinics opened in 2000, and by 2010 they numbered just over 1300. The growth of walk-in care had a great boost since then and according to the Convenient Care association, the number of them was as high as 3000 across United States in 2016 [46] .Walk-in clinics provide basic medical services such as vaccinations, evaluation of flu symptoms, and treatment of some physical injuries.

One of the major challenges of walk-in clinics is the day to day resource and staff planning because of the uncertainty in the patient demand. Patient dissatisfaction is usually a result of under planning of resources because patients wait longer periods of time to see a provider. On the other hand, if an excessive number of resources are available, a walk-in clinic will have very expensive resources idle for extended periods of time. There is also a cost associated with the coefficient of variation of service times, the probability of no-show, and the number of patients per clinical session. Authors in [22] and [23] study the problem of no-show and proposed an appointment overbooking which could reduce the negative impact of this problem but might increase the associated costs. LaGanga and Lawrence in their research work, developed a utility function for the clinic to capture the trade-off between these benefits and costs, and showed that the relative values that a clinic assigns to serving additional patients, minimizing patient waiting time [26].

The goal of this chapter is to present a decision-making model for resource planning in walk-in clinics. The model considers both management and patients' perspectives. The management perspective includes the cost of operating the clinic. The patients' perspective reflects the cost they incur as a result of waiting at the walk-in clinic[47]. For a better understanding of the proposed model, this chapter discusses a case study at a walk-in clinic located in Central Texas.

4.2 Methodology

The optimization model will be a decision-making tool to select staff capacities for the different areas in the clinic. The decision variables and the parameters for this model are in Table 11. The three model decision variables are the number of clerks at the front desk (S_1), the number of medical assistants in the assessment area (S_2), and number of providers in the examination rooms, (S_3) The objective function minimizes the total cost per hour incurred by the patients and the managers of the clinic. The first term in the objective function is a constant that represents the fixed cost incurred by the clinic to pay for the utilities and may or may not be included in the optimization model. The second term is the cost paid to the staff and providers. It is a linear combination of the decision variables and their corresponding costs in $/hour. It is well-known from queuing theory that the number of staff allocated at each service stage will affect the length of the queues. Thus, the third part of the objective function computes the total cost of waiting in queue in the clinic if Cc is the average cost of waiting in queue per hour and per patient and Lq_1 is the length of the queue at the front desk, Lq_2 is the length of the queue at the waiting room and Lq_3 is the length of the queue at the examination room.

The optimization model has three constraints. The first constraint says that the time a patient waits in a queue must not exceed a maximum value that was pre-determined after surveying management about service expectations at each areas of the clinic. The second constraint is linear and it limits the number of staff allocated by saying that the total cost per hour incurred on staff and fixed-costs must not exceed the total available budget, *b*, in $/hour. The third constraint says that the clinic must have at least one staff in each of the three service stages (front desk, assessment, and examination).

There exist closed-form expressions to calculate Lq_1, Lq_2 and Lq_3 for systems exhibiting Poisson inter-arrival times and exponential service times such as the one in this study. The reader can consult [48] or any other queueing theory book. However, these expressions are non-linear on the number of servers Sj (i.e., the decision variables in this optimization model). More importantly, the expressions require to compute a finite sum whose upper limit depends on the number of servers Sj (i.e., the decision variables in this optimization model). It makes undesirable the use of closed-form queueing theory expressions in the objective function of this optimization model. An alternative option is to model the queues as birth and death processes [48], compute the steady-state probabilities for each state and use them to find system performance measures such as the queue length, *Lq*, and the waiting time in queue, *Wq*. It is the approach followed in this work. It is amenable since the birth and death processes to model each queue and the non-linear optimization model can be set-up in Excel, linked to each other, and optimization phase can be performed with Excel Solver and the Generalized Gradient Method to solve non-linear programming problems.

a) Objective function:

$$\min TC = C_o + \sum_{j=1}^{3} S_j C_j + C_c \sum_{j=1}^{3} L_{qj} \quad [49]$$

Where,

TC = Total cost

b) Constraints:

$$(1)\ W_{qj} \leq W_{max\ j} \qquad \forall j \in \{1,2,3\}$$

$$(2)\ \sum_{j=1}^{3} S_j C_j + C_o \leq b$$

$$(3)\ S_j \geq 1$$

Table 11: Decision variables and parameters for the optimization model

Decision Variable	**Definition**
S_j	The number of staff or providers allocated to process j,where j=1 represents the front desk, j=2 the assessment room, and j=3 the waiting room
Parameters	**Definition**
W_{qj}	Average patient waiting time in queue for process j, defined by $\frac{L_{qj}}{\lambda_j}$ using Little's Law
λ_j	Average arrival rate of patients to process j
μ_j	Average service rate of process j
P_{0j}	$$\pi_j = \pi_0 c_j$$ Since at any given time, we must be in some state, the steady-state probabilities must sum to 1: $$\sum_{j=0}^{j=\infty} \pi_j = 1$$ By substituting π_j : $\pi_0 \left(1 + \sum_{j=0}^{j=\infty} C_j\right) = 1$ [48] $$\pi_0 = \frac{1}{\left(1 + \sum_{j=0}^{j=\infty} C_j\right)}$$
ρ_j	Utilization ratio $\frac{\lambda_j}{s_j \mu_j}$.
L_{qj}	Patients waiting in the queue per hour for process j.
C_o	Fixed cost for utilities such as the rent, electricity and water per hour
C_j	Cost of staff or provider at process j per hour (i.e., clerk if j=1, assessment staff if j=2, provider if j=3)
C_c	Average cost of waiting on a queue per hour.
$W_{max\ j}$	Maximum desired waiting time for a patient in queue j.
B	Total budget in $/hour

4.3 Computational Study

4.3.1 Case Study:

The walk-in clinic considered in this section is Live-Oak, located on the I-35 east access road near Wonder World Drive. It operates under the Central Texas Medical Center (CTMC), San Marcos, Texas. The clinic works from 9 a.m. to 7 p.m., Monday through Friday, and from 10 a.m. to 2 p.m. on Saturday and Sunday.

The services offered in this clinic are the treatment of colds/flu/respiratory problems, muscle/joint injuries and sprains, minor fractures, skin irritations and cuts, drug screenings and work related injuries. The clinic has a waiting room where the patients wait after they complete the check-in at the front desk. Medical assessment is the second stage of the process and it has a waiting area. The patient is called from the waiting room for tracking his/her basic vital assessment. On the third stage, patients are asked to wait in an examination room for the provider to perform the checkup.

4.3.2 Experiments:

The non-linear programming (NLP) model was built using Excel. The NLP model and the simulation models are the tools used to perform the statistical design of experiments. Three scenarios were developed. They are the morning, evening and weekend scenarios. The results of the morning scenario obtained after solving the optimization model using Excel Solver are compared with the results from the simulation model.

The design of experiments considers nine factors namely, the cost of resources and the patient's cost, arrival rate, budget, overall waiting time, maximum waiting time at the waiting room and examination room at three levels, low, medium and high. The critical factors are identified by doing an ANOVA test. The results from the simulation model are compared with the results obtained from the NLP model to identify the best combination of factors that minimizes the total costs.

Design of experiments (DOE) will be used to assess which factors affect significantly the output of the optimization model. Each one of the coefficients in the objective function will be factors in the DOE. Table 12 lists these factors and their associated costs. The data in Table 12 is based on statistics from the US department of energy, the bureau of labor statistics, and the US census department.

Table 12: Factors and associated cost

Factors	**Cost per hour**
C_o - Running cost (rent and utilities)	\$27.00
C_{j1} - Front desk staff	\$12.32
C_{j2} - Medical Assistant	\$14.48
C_{j3} – Provider	\$42.35
C_c - Patient (average pay/hour)	\$9

The waiting time at the clinic is another important response to be considered. The current waiting at clinic is more than 30 minutes at the waiting room and more than 15 minutes at the examination room.

Table 13 presents the factors included in the factorial design and its benchmark, high (10% above benchmark) and low level (10% below benchmark). The arrival rate (λj) and service rate (μj) factors were computed from the data observed. The fixed cost (Co) that includes the rent and utilities were obtained from the US department for energy, San-Marcos area. The cost associated with the resources (medical assistants and providers) (C_j) and budget per year (b) were provided by the CTMC management. The maximum waiting (Wmax j) factor is the right-hand side of the second constraint in the optimization model. Its benchmark level was assumed equal to the current waiting time at each station.

Table 13: Factorial design

		Level		
Factor	**Meaning**	**Benchmark**	**High**	**Low**
Λ	Arrival rate	5	8	3
Overall waiting (min)	sum of waiting time at both front and examination room	60		45
Wmax j1 (min)	waiting at the waiting room	42		30
Wmax j2 (min)	Waiting at the examination room	16		13
C_{j1}	Cost associated with front desk staff	$12.32	15	
C_{j2}	Cost associated with medical assistant	$14.48	18	
C_{j3}	Cost associated with provider	$42.35	50	
C_c	Cost associated with patient	$16.57	20	
B	Budget	$170	200	

4.4 ANOVA:

The ANOVA was performed using Design Expert9. As mentioned before, the goal of the experimental study is to determine the significant factors and their optimal levels. The factors considered are as follows:

The responses (R) considered in this experiment are:

- Total Cost (R1)
- Cost of clinic (R2) (i.e. cost incurred by the walk-in clinic to pay salaries to staff and providers)
- Cost for patient (R3), (i.e. total cost of waiting in the queues)
- Waiting time at the waiting room (R4)
- Waiting time at the examination room (R5)
- Front desk staff (R6)
- Medical assessment staff (R7)

- Provider (R8)

The remainder of this section presents a discussion of the results obtained for each one of these responses.

4.4.1 Total Cost: (R1)

Figure 38 and figure 39 show the half-normal plot for the total cost (R1) and the graph for the interaction between λ and b factors. Based on the half-normal plot, the significant terms for R1 are lambda (λ), budget (b), cost associated with patients (Cc), cost of medical assessment staff (Cj_2), cost of provider (Cj_3), and the interactions between lambda and budget (λ-b), lambda and cost associated with patients (λ-Cc), lambda and cost of provider (λ-Cj_3), and budget and cost of provider (b- Cj_3).

Tables 14-17 show the how R1 varies because of the changes in the levels for the interactions (λ-b), (λ-Cc), (λ-Cj_3), and (b- Cj_3) respectively.

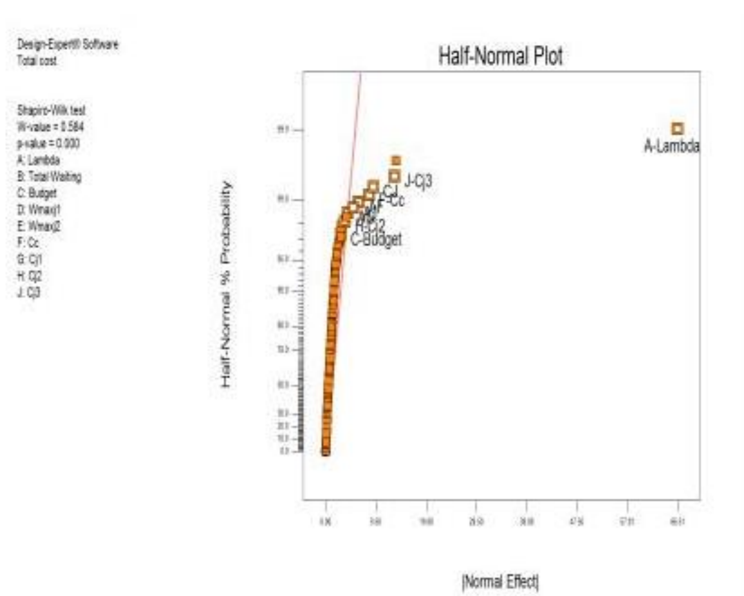

Figure 38: Half-normal plot (Total Cost)

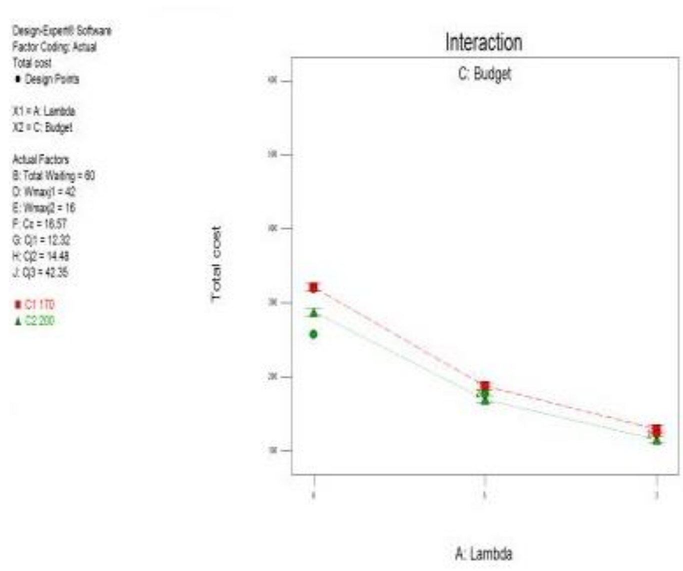

Figure 39: Interaction of significant terms (λ-b)

Table 14: Total cost vs. interaction of Lambda and Budget

Budget	**Lambda**		
	8	**5**	**3**
170	321.39	187.70	129.79
200	287.73	169.79	115.55

Table 15: Total cost vs. interaction of lambda and patient cost

Patient Cost (Cc)	**Lambda**		
	8	**5**	**3**
16.57	321.39	187.70	129.79
20	353.58	194.72	135.88

Table 16: Total cost vs. interaction of lambda and cost of provider

Cost of Provider (Cj3)	**Lambda**		
	8	**5**	**3**
42.35	321.39	187.70	129.79
50	343.18	189.76	128.11

Table 17: Total cost vs. interaction of budget and cost of provider

Cost of Provider **(Cj3)**	**Budget**	
	170	**200**
42.35	**321.39**	**287.73**
50	343.18	342.82

The tables above provide an insight about the change in the total cost due to changes in levels of the significant factors or ANOVA model terms. For instance, in Table 17, the total cost due to the interaction of budget and cost of provider is reduced by 10.47% if the budget is increased to $200 while keeping the Cj3 cost at the benchmark level (see underlined numbers in Table 17). Similarly, the interactions of other significant terms were studied.

4.4.2 Cost of clinic: (R2)

The half normal plot for the cost of clinic (R2) is shown in the figure 40. The graph for one of the significant interaction terms (λ-Cc) is given in figure 41. It is evident from figure 41 that the significant terms of R2 are lambda (λ), maximum waiting time at examination room ($Wmaxj_3$), cost associated with patient (Cc), cost of front-desk clerk (Cj_1), cost of assessment staff (Cj_3), cost of medical provider (Cj_3), and interaction of (λ-Cc), (λ-Cj_3), and (b-Cj_3).

Tables 18-20 demonstrate the change in clinic cost due to the interaction of the significant factors of the model.

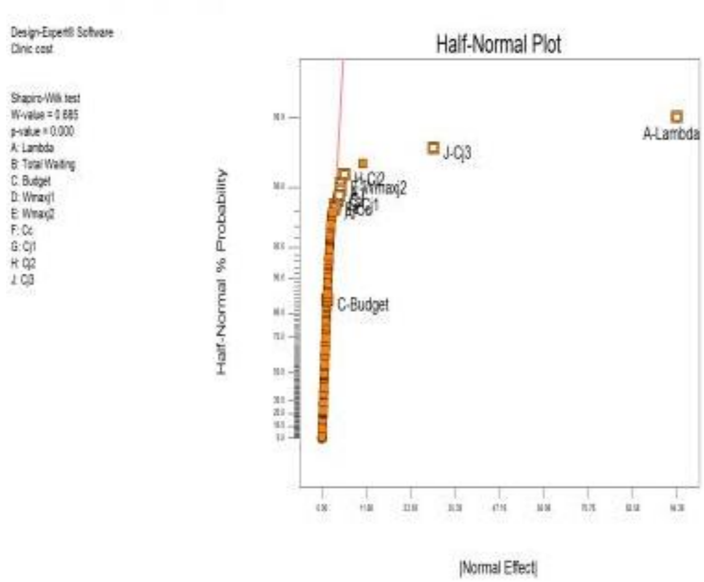

Figure 40: Half-normal plot (Clinic Cost)

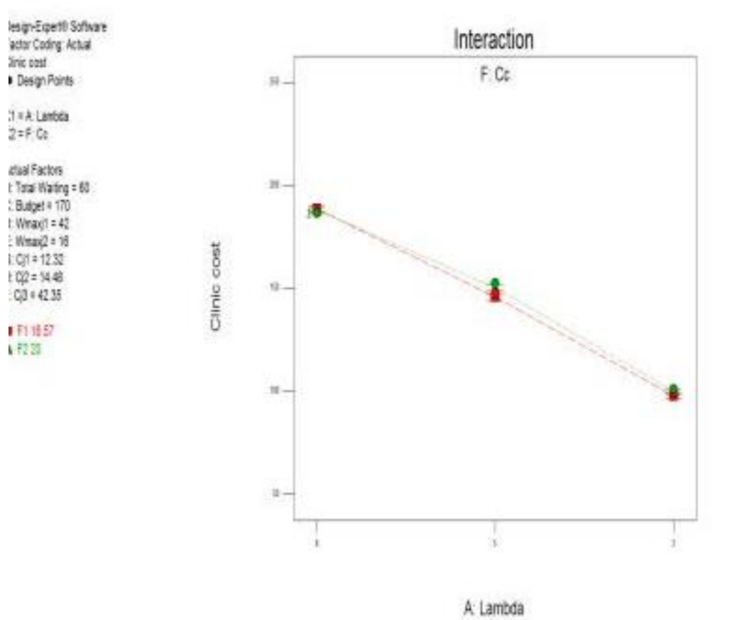

Figure 41: Interaction of significant terms (λ-Cc)

Table 18: Clinic Cost vs. interaction of lambda and patient cost

Patient Cost (Cc)	**Lambda**		
	8	**5**	**3**
16.57	188.65	145.64	97.73
20	187.98	150.23	99.68

Table 19: Clinic cost vs. interaction of lambda and cost of provider

Cost of Medical assessment staff (Cj3)	**Lambda**		
	8	**5**	**3**
42.35	188.65	145.64	97.73
50	212.36	163.97	114.40

Table 20: Clinic cost vs. interaction of budget and cost of provider

Cost of Provider (Cj3)	**Budget**	
	170	**200**
42.35	188.65	192.09
50	212.36	210.48

The interactions of the significant factors reveal that the cost of the clinic is under the budget ($200) even when patient cost, cost of medical assistant, and the cost of the provider are maintained as same as in the present system (benchmark).

4.4.3 Cost for Patient: (R3)

Response (R3), the cost of waiting in queue incurred by the patients has the factors lambda (λ), budget (b), cost associated with patients (Cc), cost of server3 (Cj3) and the interactions of lambda and budget (λ-b), lambda and cost associated with patients (λ- Cc), budget and cost of provider (b-Cj_3) as significant terms in the ANOVA model. It is shown in the half-normal plot for (R3) in the figure 42. Figure 43 show one of the significant terms of the model which is the interaction between lambda and budget.

The change in patient waiting cost (R3) based on the interaction of the factors is given in tables 21-23.

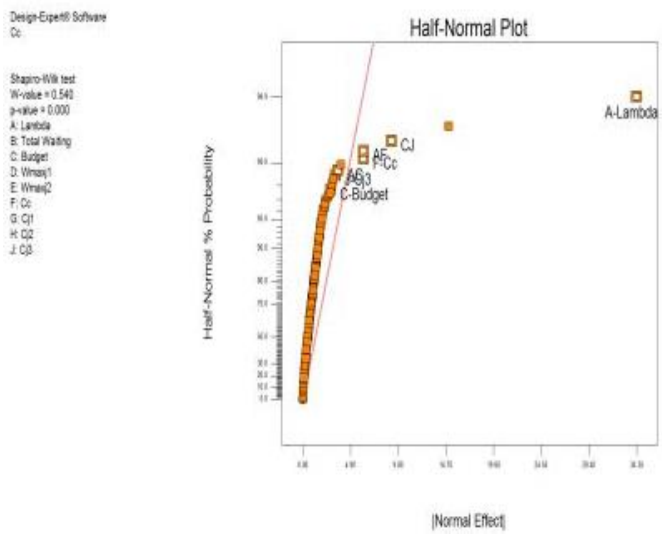

Figure 42: Half-Normal plot of patient cost

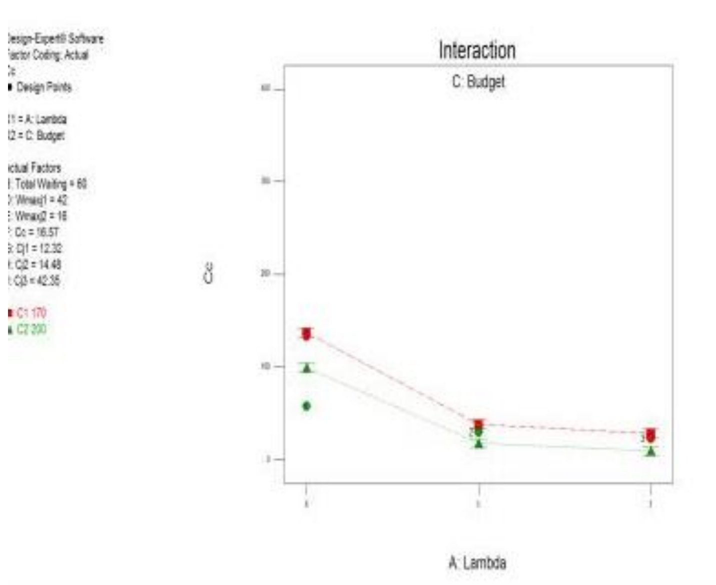

Figure 43: Patient cost vs. interaction of significant terms (λ-b)

Table 21:Interaction of lambda and budget

Budget	**Lambda**		
	8	**5**	**3**
170	99.63	18.20	9.90
200	137.06	38.48	28.36

Table 22: Patient cost vs. interaction of lambda and patient cost

Patient Cost (Cc)	**Lambda**		
	8	**5**	**3**
16.57	137.06	38.48	28.36
20	169.92	40.86	32.76

Table 23: Patient cost vs. interaction of lambda and cost of server

Cost of Provider (Cj3	**Lambda**		
	8	5	3
42.35	137.06	38.48	28.36
50	124.86	26.27	16.15

The above table reveal that the cost associated with patients varies mainly based on the arrival rate of the patients (lambda) and its interaction with other significant factors.

4.4.4 Waiting time at the front desk (W1): (R4)

The waiting time at the front desk (R4) has lambda (λ), maximum Waiting time (TW) a customer will experience at the clinic, budget (b), cost of server1 (Cj_1), cost of server3 (Cj_3) and the interactions (λ- TW), (λ- b), (λ- Cj_1), (λ- Cj_3), and (b- Cj_3) as the significant terms in the ANOVA model. It is demonstrated in figure 44 by the half-normal plot for ($Wmaxj_1$). Figure 45 show the behavior of the interaction between λ and TW.

Tables 24-28 show the maximum waiting time at the front desk during the interaction of (λ- TW), (λ- b), (λ- Cj1), (λ- Cj_3), and (b- Cj_3).

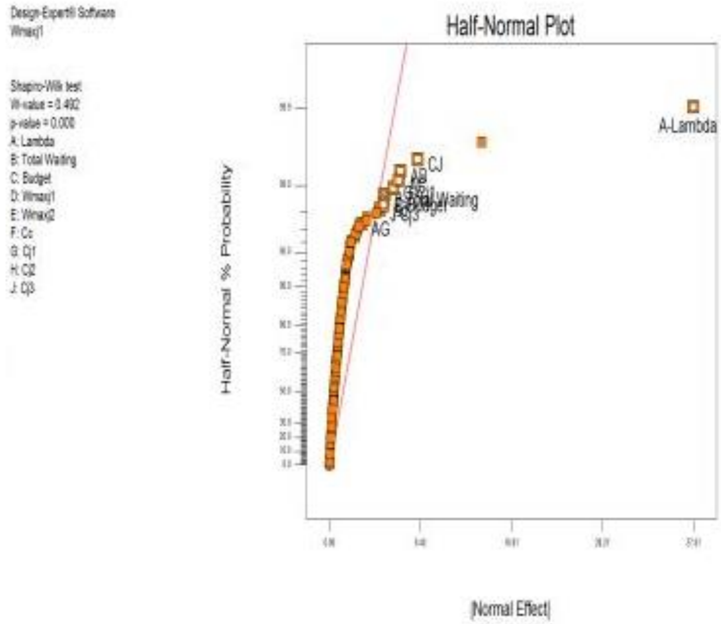

Figure 44: Half-normal plot of waiting time at front desk

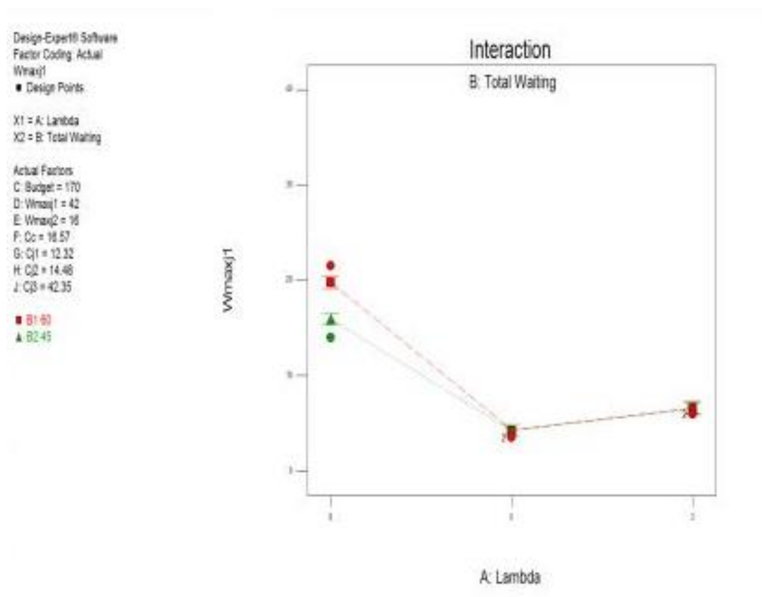

Figure 45: Interaction of significant terms (λ-overall waiting)

Table 24: Waiting time at the front desk (minutes) vs. interaction of lambda and overall waiting

Overall Waiting	**Lambda**		
	8	**5**	**3**
60	20.82	3.81	6.59
45	19.51	2.50	5.27

Table 25: Waiting time at the front desk (minutes) vs. interaction of lambda and budget

Budget	**Lambda**		
	8	**5**	**3**
170	20.82	3.81	6.59
200	10.45	3.92	6.59

Table 26: Waiting time at the front desk (minutes) vs. interaction of lambda and cost of Front desk staff

Cost of Front desk staff (Cj1)	**Lambda**		
	8	**5**	**3**
12.32	20.82	3.81	6.59
15	23.57	4.86	7.39

Table 27: Waiting time at the front desk (minutes) vs. interaction of lambda and cost of provider

Cost of provider (Cj3)	**Lambda**		
	8	**5**	**3**
42.35	20.82	3.81	6.59
50	17.34	4.57	6.59

Table 28: Waiting time at the front desk (minutes) vs. interaction of budget and cost of provider

Cost of provider (Cj3)	Budget	
	170	**200**
42.35	20.82	10.45
50	17.34	20.23

The waiting time for assessment can be reduced by 73.2% if the budget increases to $200. Currently the waiting time is more than 30 minutes, and it can go to 10.45 minutes as seeing on Table 28. Similarly, the patient flow also has an impact on the waiting time at the front desk. For instance, table 25 show the values of the waiting time at different levels of budget and the arrival rate of patients. By studying these interactions, combination of factors that yields the shortest waiting time as well as low cost can be selected.

4.4.5 Waiting time at examination room (W2): (R5)

Figure 46 show the half-normal plot for the maximum waiting time at the examination room. It demonstrates the significant factors for (R5) waiting time at the examination room (W2) in the ANOVA model which are lambda (λ), waiting time at examination room ($Wmaxj_2$), and the interaction between budget and the cost of provider (b- Cj_3). The interaction between budget and cost of provider is shown in the figure 47.

Table 29 gives the maximum waiting time at the examination room when budget and cost of provider interact.

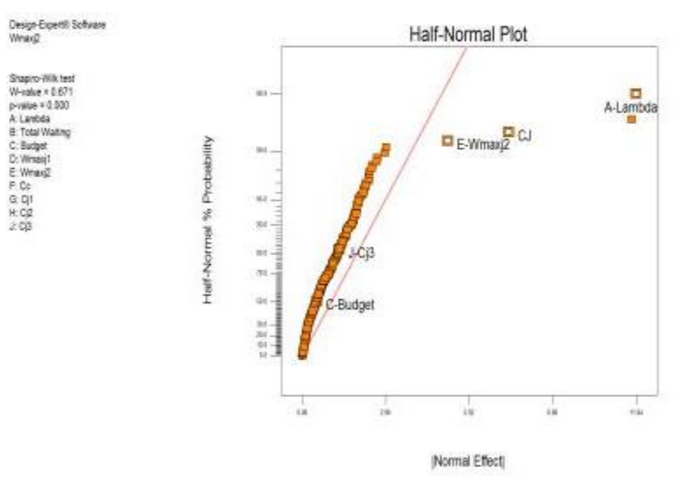

Figure 46: Half-normal plot of waiting time at examination room

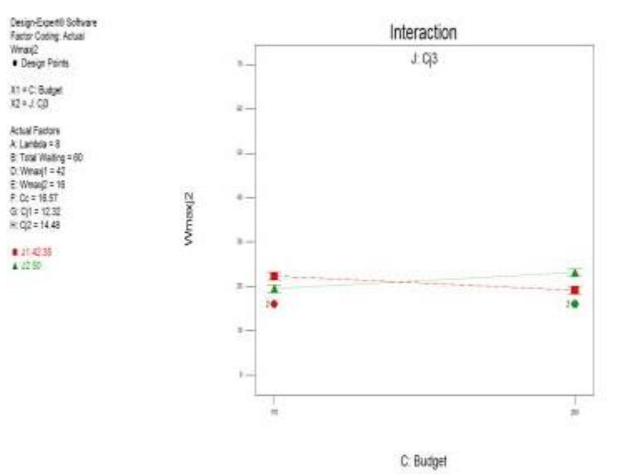

Figure 47: Interaction of significant terms (b-Cj3)

Table 29: Waiting time at the examination room (minutes) vs. interaction of budget and cost of provider

Cost of provider (Cj3)	**Budget**	
	170	**200**
42.35	22.27	19.16
50	19.52	23.10

The waiting time at the examination room depends on the cost associated with the doctors and the budget. The results from the Table 29 paves way to determine the number of servers at the examination room based on the cost factor as well as the waiting time. For example, the waiting time is 13.96% less when the cost of provider is $42.35 and budget is set at $200 compared to when the budget is $170. Because after the budget increased, the optimization model could increase the number of providers which resulted in reduced waiting time.

4.4.6 Number of clerks at the front desk (R6)

The half-normal plot for the response (R6) (Front desk staff) in figure 48 show that it has significant interactions such as (TW-b), (b-Cj1), (Wmaxj1-Cc), (TW-Cj2-Cj3), (b-Wmaxj1-Cc), (Cj1-Cj2-Cj3), (TW-b-Cj2-Cj3), (TW-Wmaxj1-Cc-Cj1), (b-Cj1-Cj2-Cj3), (Wmaxj1-Cc-Cj2-Cj3), (TW-b-Wmaxj1-Cc-Cj1), and (b-Wmaxj1-Cc-Cj2-Cj3), (TW-Wmaxj1-Cc-Cj1-Cj2-Cj3), and (TW-b-Wmaxj1-Cc-Cj1-Cj2-Cj3). The significant factors are lambda (λ), overall waiting time (TW) and cost of Front desk staff (Cj_1). Figure 49 show the interaction of TW and b.

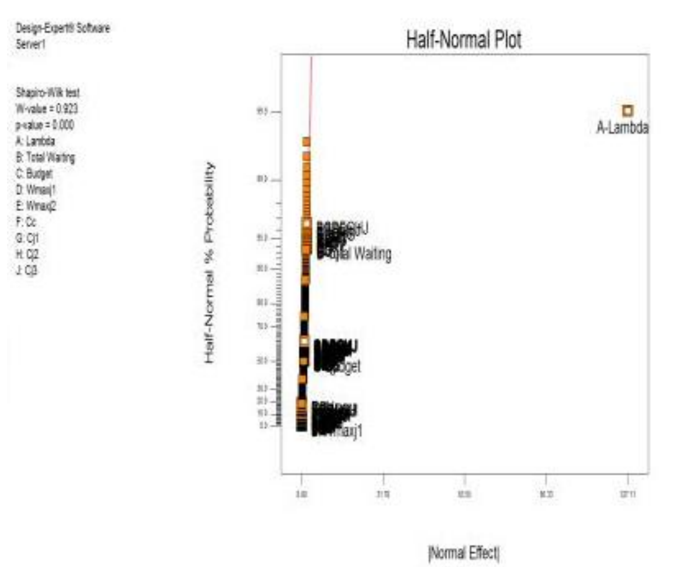

Figure 48: Half-normal plot for Front desk staff

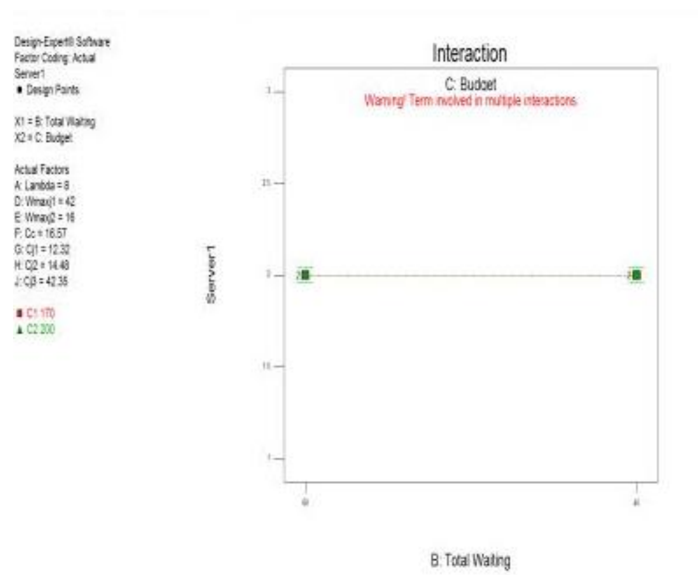

Figure 49: Interaction of significant terms (overall waiting-budget)

After observing the behavior of the response variable R6 to the significant interactions the optimal number of servers at the front desk is always 2.

4.4.7 Number of medical assessment staff: (R7)

Response (R7), number of medical assessment staff have significant factors such as lambda (λ), overall waiting time (TW), budget (b), cost associated with patients (Cc), cost of server2 (Cj2), cost of server3 (Cj3) and significant interaction of factors such as (TW-b), (λ- TW), (λ- b), (λ- Cj_3), (b- Cj_3), (λ- Cc), (λ- Cj_2), (Cc- Cj_2) . The significant terms are shown in the half-normal plot which is demonstrated in the figure 50. Figure 51 demonstrates one of the interaction of significant factors (λ- TW).

Tables 30-37 provide information about the number of servers at the assessment room that result from the of interaction between some of the significant model terms.

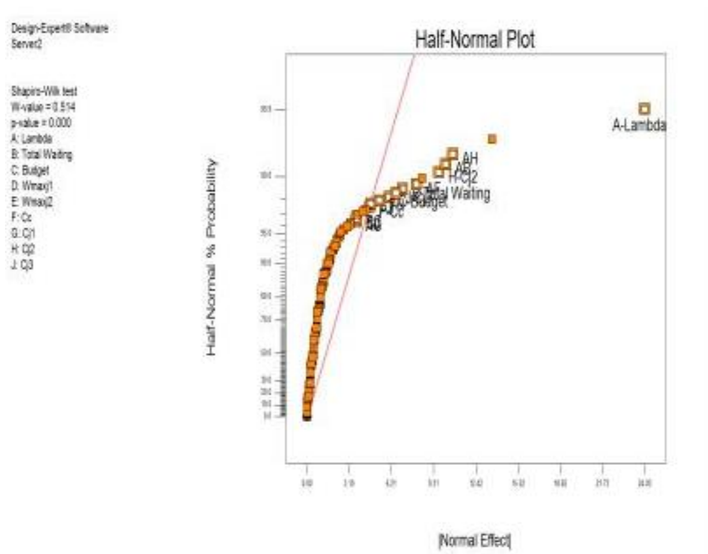

Figure 50: Half-normal plot for medical assessment staff

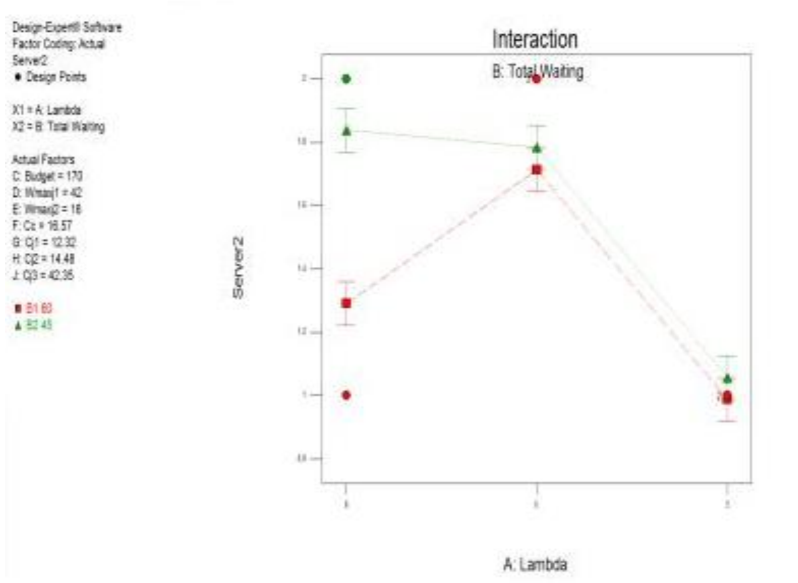

Figure 51: Interaction of significant terms (overall waiting-lambda)

Table 30: Staff in the assessment room vs. interaction of lambda and overall waiting

Overall waiting	**Lambda**		
	8	**5**	**3**
60	1	2	1
45	2	2	1

Table 31: Staff in the assessment room vs. interaction of lambda and budget

Budget	**Lambda**		
	8	**5**	**3**
170	1	2	1
200	2	2	1

Table 32: Staff in the assessment room vs. interaction of lambda and patient cost

Patient Cost (Cc)	**Lambda**		
	8	5	3
16.57	1	2	1
20	1	2	1

Table 33: Staff in the assessment room vs. interaction of lambda and cost of medical assessment staff

Cost of Front desk staff (Cj1)	**Lambda**		
	8	**5**	**3**
14.48	1	2	1
18	1	1	1

Table 34: Staff in the assessment room vs. interaction of lambda and cost of provider

Cost of Front desk staff (Cj1)	**Lambda**		
	8	**5**	**3**
42.35	1	2	1
50	1	2	1

Table 35: Staff in the assessment room vs. interaction of budget and overall waiting

Budget	**Overall waiting**	
	60	**45**
170	1	2
200	2	2

Table 36: Staff in the assessment room vs. interaction of budget and cost of provider

Cost of provider (Cj3)	**Budget**	
	170	**200**
42.35	1	2
50	1	1

Table 37: Staff in the assessment room vs. interaction of patient cost and cost of medical assessment staff

Cost of medical assessment staff (Cj2)	**Patient Cost**	
	16.57	**20**
14.48	1	1
18	1	1

4.4.8 Number of providers: (R8)

Figure 52 provide the significant model terms for the response R8 (number of medical providers) through the half normal plot that is shown in Figure 52. The significant terms in the ANOVA model are lambda (λ), waiting time at examination room ($Wmaxj_2$), and the interaction of lambda and waiting time at the examination room (λ-$Wmaxj_2$, figure 53).

Table 38 shows the number of medical providers at the examination room vs. the interaction of lambda and maximum waiting time at the examination room ($Wmaxj_2$).

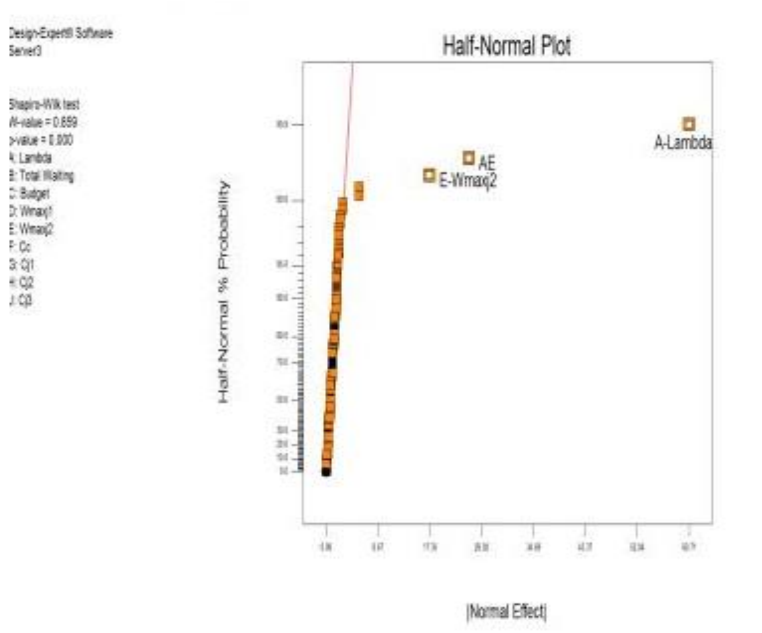

Figure 52: Half-normal plot for provider

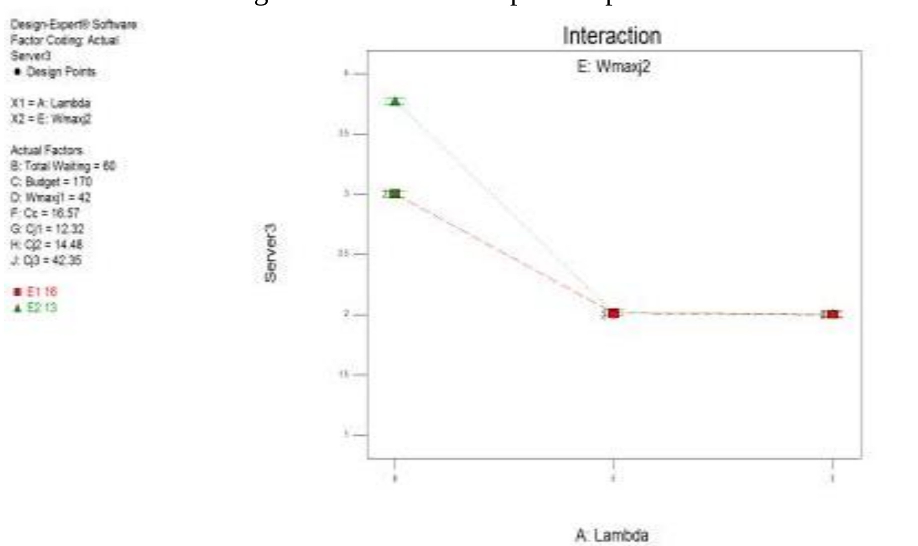

Figure 53: Interaction of significant terms (lambda-waiting time at examination room)

Table 38: Number of medical providers vs. interaction of lambda and desired or maximum waiting time allowed at the examination room

$Wmaxj_2$	**Lambda**		
	8	**5**	**3**
16	3	2	2
13	4	2	2

The number of providers to be allocated in the walk-in clinic can be decided based on the target waiting time (less than 30 minutes) to be achieved and the different levels of arrival rate. The results from table 38 reveal that, to reduce the waiting time to 16 minutes at the examination room, there must be two providers during normal and low patient flow, whereas three providers during high flow of patients. However, by taking other factors such as the arrival rate, budget, total waiting time, cost of servers, and cost for patients into consideration and by comparing the results of other responses, the total number of server at the examination room can be increased to either two or more based on the constraints.

4.5 Results

The experiments were performed by considering nine factors in which the factor lambda (arrival rate) had three levels and the other eight factors mentioned in section 4.4 had two levels. A total of 768 runs were performed and the values were computed for eight responses. The experiments were performed by using Excel solver and the analysis of the computed values was performed using Design Expert9.

Based on the experiments performed and analysis of the results, a trade-off between the total cost in the clinic and the cost associated with the patient is achieved by selecting the best combination of servers at each station and by reducing the waiting time under low, normal, and high demand. The results were sorted in an ascending manner and for each arrival rate λ (low, normal, high), all the combinations that satisfied the constraints with minimum overall cost were selected.

Tables 39-41 provide the factors and their corresponding responses that yield the most efficient responses at low, normal, and high patient flow respectively.

Table 39: Low arrival rate – Factors and responses

λ	TW	b	Wmaxj1	Wmaxj2?	(Cc)	(Cj1)	(Cj2)	(Cj3)	TC	Cost of clinic	Cc	W1	W2	Server		
3	**45**	**170**	**30**	**13**	**16.57**	**12.32**	**14.48**	**50**	$118.05	$97.07	$21	5.95	13.00	1	1	1
3	**60**	**170**	**30**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$122.96	$99.50	$23	5.95	16.00	1	1	2
3	**60**	**200**	**42**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$122.96	$99.50	$23	5.95	16.00	1	1	2
3	**45**	**200**	**42**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$122.96	$99.50	$23	5.95	16.00	1	1	2

Table 40: Normal arrival rate – Factors and responses

λ	TW	B	Wmax j1	Wmax j2?	(Cc)	(Cj1)	(Cj2)	(Cj3)	TC	Cost of clinic	Cc	W1	W2	Server		
5	**60**	**170**	**42**	**13**	**16.57**	**12.32**	**14.48**	**42.35**	$ 165.59	$149.02	$17	3.53	13.00	2	2	2
5	**45**	**170**	**30**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$ 177.67	$147.22	$30	3.53	14.23	2	2	2
5	**60**	**200**	**42**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$ 177.67	$147.22	$30	3.53	14.23	2	2	2
5	**45**	**200**	**42**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$ 177.67	$147.22	$30	3.53	14.23	2	2	2

Table 41: High arrival rate – Factors and responses

λ	TW	B	Wmaxj1	Wmaxj3	(Cc)	(Cj1)	(Cj2)	(Cj3)	TC	Cost of clinic	Cc	W1	W2	Server		
8	**60**	**170**	**30**	**13**	**16.57**	**15**	**14.48**	**50**	$ 207.27	$172.07	$35	3.56	13.99	2	2	2
8	**60**	**200**	**30**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$ 257.74	$200.00	$58	5.30	16.00	2	2	3
8	**45**	**200**	**42**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$ 257.74	$200.00	$58	5.30	16.00	2	2	3
8	**45**	**170**	**30**	**13**	**16.57**	**15**	**14.48**	**42.35**	$ 259.34	$200.10	$39	3.62	10.68	2	2	4

The optimal combination of servers that resulted in a trade-off between total cost of the clinic and cost associated with patient was selected by comparing the values from the above tables. For instance, during a normal patient flow (λ = 5), by increasing the providers' capacity to 2 (H) the waiting time is reduced by 90% at the waiting room and 14% at the examination room while keeping the cost of the clinic 26% less than the budget. Similarly, during high patient flow it can be seen from table 42 that the waiting time at the waiting area is reduced by 87% while the clinical cost is equal to the budget allotted.

Table 42: Optimal values

λ	TW	b	Wmaxj1	Wmaxj3	(Cc)	(Cj1)	(Cj2)	(Cj3)	TC	Cost of clinic	Cc	W1	W2	Server		
3	**45**	**200**	**42**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$122.96	$99.50	$23	5.95	13.00	1	1	2
5	**45**	**200**	**42**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$177.67	$147.22	$30	3.53	14.23	2	2	2
8	**45**	**200**	**42**	**16**	**16.57**	**12.32**	**14.48**	**42.35**	$257.74	$200.00	$58	5.30	16.00	2	2	3

Therefore, based on the study made on the daily patient flow for over a period of six months and by observing the responses mentioned in section 5.4. By maintaining the server at front desk as one, at assessment room as two and by increasing the number of provider at examination room as two will result in an efficient system which has a trade-off between the cost of clinic and, the overall waiting time and length of stay.

5. CONCLUSION

Patient satisfaction and quality of service are of paramount importance for the operation of a walk-in clinic. This work investigated the allocation of resources, length of stay of patient in the system and cost associated with each process, from the clinic and the patients' perspective. While there is extensive literature that investigates the scheduling of appointments and the resource allocation problems in outpatient clinics, open access clinics, and emergency rooms, the operations at walk-in clinics have not been previously studied with regards to proper resource allocation and patient satisfaction. This thesis makes a significant contribution to the health care management literature concerned with improving the clinic operations and the decision-making policies in health care.

A simulation model was built to find the right combination of resources to be allocated at each station in the clinic. First, data was collected from the clinic to build a simulation model to simulate the current scenario at the clinic. Secondly, by design of experiments, a set of 27 experiments were performed and simulation models were built for each scenario considered (morning, afternoon, and weekend). Based on the results, an efficient combination was identified and the model was validated and verified. As discussed in section 3.6, the results from the simulation model showed that by increasing the provider capacity to two (H), there is a 38% reduction in the waiting time between experiments 10 and 15 for the morning scenario. Similarly, a 70% and 64% improvement is observed in the afternoon and weekend scenario, respectively. The cycle time was reduced by 34% by increasing the provider capacity (experiment 15, **NHN**) to 2 (H) and during high patient flow as shown in experiment 26 **(NHH**), having two providers (H) reduced the cycle time by 11% compared to actual system.

A nonlinear-programming model was used to support the results of simulation model. The model finds a trade-off between the cost and the system efficiency to increase the patient satisfaction with good quality of care. A total 768 experiments were performed after considering nine factors at normal, high, and low levels. The NLP model was built using Excel solver and the results from the experiments revealed the optimal number of resources at each station and the cost associated with them. The total clinic expense, cost with respect to patients and the overall waiting time were some of the experimental responses.

The results from the NLP model revealed that by increasing the provider capacity, the waiting time is reduced by 90% at the waiting room and 14% at the examination room while keeping the cost of the clinic 26% less than the clinic's budget during normal patient flow. Similarly, during high patient flow (H) it can be seen from table 43 that the waiting time at waiting area is reduced by 87% while the clinical cost is equal to the budget allotted by the clinic management.

The analysis of data obtained from the Live-Oak walk-in care demonstrate that having one server at the front desk, two medical assistants and two providers form an efficient system without increase in the total cost and with high quality of service. This is evident from the output of the simulation and analytical models. The results from NLP model are demonstrated in the figure 54 and figure 55.

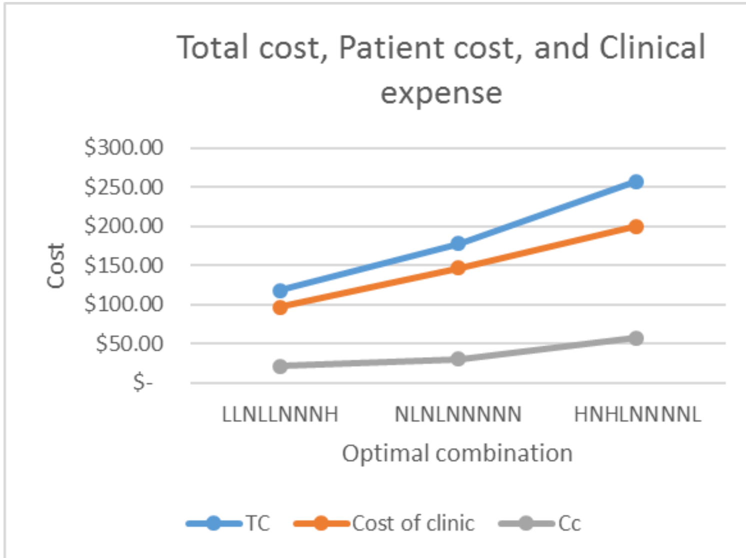

Figure 54: Total cost, clinic cost, and patient cost at the optimal combination of staff and medical providers

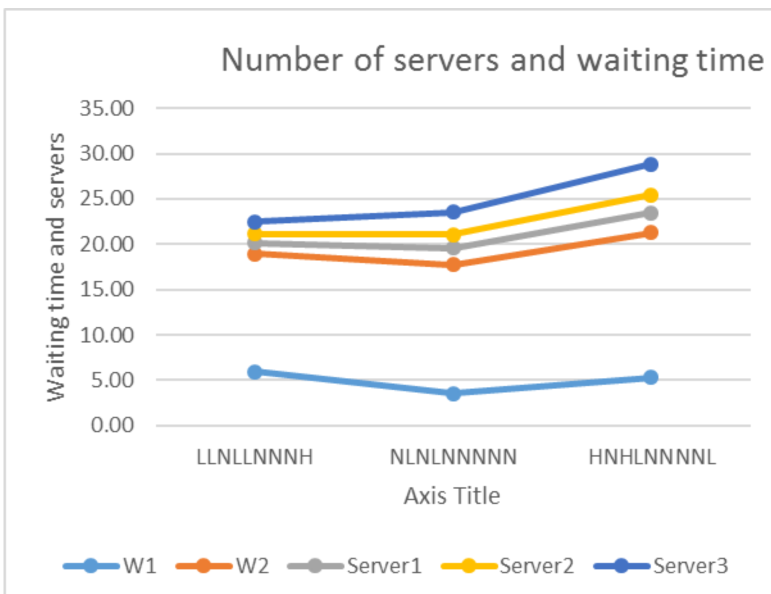

Figure 55: Waiting time at front desk and examination room with the optimal number of staff/medical providers at each station

The activities outlined in this research are *novel*, while engaging with a committed, well-respected clinical site. Implementing the optimal results observed from the optimization and the simulation models will lead to increase in the number of patients served in the working hours, reduce patient waiting time, and improve quality of care and customer satisfaction.

The research results provide important data and information for health care organizations and, perhaps ultimately enable policy makers to recommend more efficient intervention strategies. These strategies will impact the health care industry in Texas, across the nation, and may also have international implications.

REFERENCES

[1] M. A. Kallen, J. A. Terrell, P. Lewis-Patterson, and J. P. Hwang, "Improving wait time for chemotherapy in an outpatient clinic at a comprehensive cancer center," *Journal of Oncology Practice,* vol. 8, pp. e1-e7, 2012.

[2] Y.-L. Huang, W. M. Hancock, and G. D. Herrin, "An alternative outpatient scheduling system: Improving the outpatient experience," *IIE Transactions on Healthcare Systems Engineering,* vol. 2, pp. 97-111, 2012.

[3] H. Zhu, M. Hou, C. Wang, and M. Zhou, "An efficient outpatient scheduling approach," *IEEE Transactions on Automation science and engineering,* vol. 9, pp. 701-709, 2012.

[4] T. Cayirli and K. K. Yang, "A Universal Appointment Rule with Patient Classification for Service Times, No-Shows, and Walk-Ins," *Service Science,* vol. 6, pp. 274-295, 2014.

[5] J. Tang, C. Yan, and P. Cao, "Appointment scheduling algorithm considering routine and urgent patients," *Expert Systems with Applications,* vol. 41, pp. 4529-4541, 2014.

[6] N. Liu, S. Ziya, and V. G. Kulkarni, "Dynamic scheduling of outpatient appointments under patient no-shows and cancellations," *Manufacturing & Service Operations Management,* vol. 12, pp. 347-364, 2010.

[7] P.-F. J. Tsai and G.-Y. Teng, "A stochastic appointment scheduling system on multiple resources with dynamic call-in sequence and patient no-shows for an outpatient clinic," *European Journal of Operational Research,* vol. 239, pp. 427-436, 2014.

[8] 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, pp. 1211-1219, 2006.

[9] J. Patrick, "A Markov decision model for determining optimal outpatient scheduling," *Health Care Manag Sci,* vol. 15, pp. 91-102, Jun 2012.

[10] J. Feldman, N. Liu, H. Topaloglu, and S. Ziya, "Appointment Scheduling Under Patient Preference and No-Show Behavior," *Operations Research,* vol. 62, pp. 794-811, 2014.

[11] B. Zeng, A. Turkcan, J. Lin, and M. Lawley, "Clinic scheduling models with overbooking for patients with heterogeneous no-show probabilities," *Annals of Operations Research,* vol. 178, pp. 121-144, 2009.

[12] L. R. LaGanga, "Lean service operations: Reflections and new directions for capacity expansion in outpatient clinics," *Journal of Operations Management,* vol. 29, pp. 422-433, 2011.

[13] B. Mocarzel, D. Shelton, B. Uyan, E. Pérez, J. A. Jimenez, and L. DePagter, "Modeling and simulation of patient admission services in a multi-specialty outpatient clinic," in *Proceedings of the 2013 Winter Simulation Conference: Simulation: Making Decisions in a Complex World,* 2013, pp. 2309-2319.

[14] T. Sowle, N. Gardini, F. V. A. Vazquez, E. Pérez, J. A. Jimenez, and L. DePagter, "A simulation-IP based tool for patient admission services in a multi-specialty outpatient clinic," in *Proceedings of the 2014 Winter Simulation Conference*, 2014, pp. 1186-1197.

[15] D. Walker, E. Shanks, D. Montoya, C. Weiman, E. Pérez, and L. DePagter, "Towards a simulation based methodology for scheduling patient and providers at outpatient clinics," in *Proceedings of the 2015 Winter Simulation Conference*, 2015, pp. 1515-1524.

[16] J. Tang, C. Yan, and R. Y. K. Fung, "Optimal appointment scheduling with no-shows and exponential service time considering overtime work," *Journal of Management Analytics,* vol. 1, pp. 99-129, 2014.

[17] C.-J. Ho and H.-S. Lau, "Minimizing total cost in scheduling outpatient appointments," *Management science,* vol. 38, pp. 1750-1764, 1992.

[18] T. Cayirli and E. D. Gunes, "Outpatient appointment scheduling in presence of seasonal walk-ins," *Journal of the Operational Research Society,* vol. 65, pp. 512-531, 2013.

[19] M. Guo, M. Wagner, and C. West, "Outpatient clinic scheduling-a simulation approach," in *Simulation Conference, 2004. Proceedings of the 2004 Winter*, 2004, pp. 1981-1987.

[20] R. Kopach, P.-C. DeLaurentis, M. Lawley, K. Muthuraman, L. Ozsen, R. Rardin, *et al.*, "Effects of clinical characteristics on successful open access scheduling," *Health Care Management Science,* vol. 10, pp. 111-124, 2007.

[21] L. W. Robinson and R. R. Chen, "A Comparison of Traditional and Open-Access Policies for Appointment Scheduling," *Manufacturing & Service Operations Management,* vol. 12, pp. 330-346, 2010.

[22] L. W. Robinson and R. R. Chen, "The Effects of Patient No-Shows on Traditional and Open-Access Appointment Scheduling Policies," *SSRN Electronic Journal*, 2009.

[23] Y. Peng, X. Qu, and J. Shi, "A hybrid simulation and genetic algorithm approach to determine the optimal scheduling templates for open access clinics admitting walk-in patients," *Computers & Industrial Engineering,* vol. 72, pp. 282-296, 2014.

[24] X. Qu, R. L. Rardin, and J. A. S. Williams, "Single versus hybrid time horizons for open access scheduling," *Computers & Industrial Engineering,* vol. 60, pp. 56-65, 2011.

[25] L. R. LaGanga and S. R. Lawrence, "Appointment Overbooking in Health Care Clinics to Improve Patient Service and Clinic Performance," *Production and Operations Management,* vol. 21, pp. 874-888, 2012.

[26] L. R. LaGanga and S. R. Lawrence, "Clinic overbooking to improve patient access and increase provider productivity," *Decision Sciences*, vol. 38, pp. 251-276, 2007.

[27] X. Qu, R. L. Rardin, J. A. S. Williams, and D. R. Willis, "Matching daily healthcare provider capacity to demand in advanced access scheduling systems," *European Journal of Operational Research,* vol. 183, pp. 812-826, 2007.

[28] X. Qu, R. L. Rardin, and J. A. S. Williams, "A mean–variance model to optimize the fixed versus open appointment percentages in open access scheduling systems," *Decision Support Systems,* vol. 53, pp. 554-564, 2012.

[29] S. Lee and Y. Yih, "Analysis of an Open Access Scheduling System in Outpatient Clinics: A Simulation Study," *Simulation,* vol. 86, pp. 503-518, 2010.
[30] K. D. Rose, J. S. Ross, and L. I. Horwitz, "Advanced access scheduling outcomes: a systematic review," *Arch Intern Med*, vol. 171, pp. 1150-9, Jul 11 2011.
[31] X. Qu and J. Shi, "Modeling the effect of patient choice on the performance of open access scheduling," *International Journal of Production Economics,* vol. 129, pp. 314-327, 2011.
[32] J. Lin, K. Muthuraman, and M. Lawley, "Optimal and approximate algorithms for sequential clinical scheduling with no-shows," *IIE Transactions on Healthcare Systems Engineering,* vol. 1, pp. 20-36, 2011/03/16 2011.
[33] S. Chakraborty, K. Muthuraman, and M. Lawley, "Sequential clinical scheduling with patient no-shows and general service time distributions," *IIE Transactions,* vol. 42, pp. 354-366, 2010.
[34] B. P. Morrison and B. C. Bird, "Healthcare process analysis: a methodology for modeling front office and patient care processes in ambulatory health care," in *Proceedings of the 35th conference on Winter simulation: driving innovation*, 2003, pp. 1882-1886.
[35] C. Duguay and F. Chetouane, "Modeling and improving emergency department systems using discrete event simulation," *Simulation-Transactions of the Society for Modeling and Simulation International*, vol. 83, pp. 311-320, Apr 2007.
[36] D. Gupta and B. Denton, "Appointment scheduling in health care: Challenges and opportunities," *IIE Transactions,* vol. 40, pp. 800-819, 2008.
[37] D. A. Thompson, P. R. Yarnold, D. R. Williams, and S. L. Adams, "Effects of actual waiting time, perceived waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department," *Annals of emergency medicine,* vol. 28, pp. 657-665, 1996.
[38] J. C. Mowen, J. W. Licata, and J. McPhail, "Waiting in the emergency room: how to improve patient satisfaction," *Marketing Health Services,* vol. 13, p. 26, 1993.
[39] D. W. Spaite, F. Bartholomeaux, J. Guisto, E. Lindberg, B. Hull, A. Eyherabide, *et al.*, "Rapid process redesign in a university-based emergency department: Decreasing waiting time intervals and improving patient satisfaction," *Annals of Emergency Medicine,* vol. 39, pp. 168-177, 2002.
[40] N. Mandahawi, S. Al-Shihabi, A. A. Abdallah, and Y. M. Alfarah, "Reducing waiting time at an emergency department using design for Six Sigma and discrete event simulation," *International Journal of Six Sigma and Competitive Advantage,* vol. 6, pp. 91-104, 2010.
[41] P. Jones and K. Schimanski, "The four hour target to reduce Emergency Department 'waiting time': a systematic review of clinical outcomes," *Emerg Med Australas,* vol. 22, pp. 391-8, Oct 2010.
[42] Y. B. Ferrand, M. J. Magazine, and U. S. Rao, "Managing operating room efficiency and responsiveness for emergency and elective surgeries—A literature survey," *IIE Transactions on Healthcare Systems Engineering,* vol. 4, pp. 49-64, 2014/01/02 2014.
[43] M. Lamiri, X. Xie, A. Dolgui, and F. Grimaud, "A stochastic model for operating room planning with elective and emergency demand for surgery," *European Journal of Operational Research,* vol. 185, pp. 1026-1037, 2008.

[44] N. Ayvaz and W. T. Huh, "Allocation of hospital capacity to multiple types of patients," *Journal of Revenue and Pricing Management*, vol. 9, pp. 386-398, 2010.

[45] G. Latorre-Núñez, A. Lüer-Villagra, V. Marianov, C. Obreque, F. Ramis, and L. Neriz, "Scheduling operating rooms with consideration of all resources, post anesthesia beds and emergency surgeries," *Computers & Industrial Engineering*, vol. 97, pp. 248-257, 2016.

[46] C. E. Pollack, C. Gidengil, and A. Mehrotra, "The growth of retail clinics and the medical home: two trends in concert or in conflict?," *Health Affairs*, vol. 29, pp. 998-1003, 2010.

[47] K. R. Yabroff, G. P. Guy Jr, D. U. Ekwueme, T. McNeel, H. M. Rozjabek, E. Dowling, *et al.*, "Annual patient time costs associated with medical care among cancer survivors in the United States," *Medical care*, vol. 52, p. 594, 2014.

[48] W. L. Winston and J. B. Goldberg, *Operations research: applications and algorithms* vol. 3: Duxbury press Boston, 2004.

[49] H. Mehri, T. Djemel, and H. Kammoun, "SOLVING OF WAITING LINES MODELS IN THE AIRPORT USING QUEUING THEORY MODEL AND LINEAR PROGRAMMING THE PRACTICE CASE: AIMHB," 2006.