KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY

Modelling Queuing system in Healthcare centres. A case study of the dental department of the Essikado Hospital, Sekondi

By
Rebecca Nduba Quarm

A THESIS SUBMITTED TO THE DEPARTMENT OF MATHEMATICS, KWAME NKRUMAH UNIVERSITY OF SCIENCE AND TECHNOLOGY IN PARTIAL FUFILLMENT OF THE REQUIREMENT FOR THE DEGREE OF M.PHIL APPLIED MATHEMATICS

October 20, 2015

Declaration

I hereby declare that this submission is my own work towards the award of the M. Phil degree and that, to the best of my knowledge, it contains no material previously published by another person nor material which had been accepted for the award of any other degree of the university, except where due acknowledgement had been made in the text.

Rebecca Nduba Quarm(PG 1100013)	....	....
Student	Signature	Date

Certified by:

Mr. K.F Darkwah	....	....
Supervisor	Signature	Date

Certified by:

Prof. S.K. Amponsah	....	....
Head of Department	Signature	Date

Dedication

This work is dedicated to my husband, Joel, and my lovely children: Johannah, Daniel and Genevieve

Abstract

In this study, the queuing characteristics at the dental department of Essikado Hospital in the Sekondi-Takoradi Metropolis was analysed using a Single-server queuing Model. The data was collected over a period of one week, from 8:30 am-11:30 am each day. By using the queuing rule First-come, First-served as practiced by the case study and M/M/s queuing model, the performance measures were calculated. The average mean arrival and mean service times were found to be 1.6 patients/hour and 4.4477 patients/hour respectively. The average number of patients in the system and in the queue was 0.5619 patients and 0.2021 patients respectively. Also, the average time spent in the system and the average time it takes for service to start was 0.3512 hours and 0.1263 hours respectively. Recommendations made include obtaining a central tray setup system for instruments required for the different dental procedures done in the dental department so as to reduce the time spent sorting them up thereby reducing the amount of time spent in queue and also increasing the number of dentist from one to two.

Contents

List of Tables

List of Figures

Chapter 1

INTRODUCTION

1.1 Overview

Queuing theory deals with the study of queues which abound in practical situations and arise so long as arrival rate of any system is faster than the system can handle. Following Nkeiruka et al (2013), queuing theory is applicable to any situation in general life ranging from customers arriving at a bank, customers at a supermarket waiting to be attended to by a cashier and in health care settings. Queuing theory is the mathematical approach to the analysis of waiting lines in any setting where arrival rate of subjects is faster than the system can handle. In queuing theory a model is constructed so that queue lengths and waiting times can be predicted (Sundarapandian, 2009). In situations where facilities are limited and cannot satisfy the demand made upon them, bottlenecks occur which manifest as queue but patients are not interested in waiting in queues. When patients are not interested in waiting in queues, there is danger that waiting time will become excessive leading to the loss of some patients to competitors.

Customer satisfaction has become a serious concern in the service sector. In the healthcare industry, a number of initiatives have been introduced to enhance customer satisfaction. The healthcare industry providers globally are experiencing increasing pressure to concurrently reduce cost and improve the access and quality of care they deliver. Many healthcare institutions are confronted with long waiting times, delays and queues of patients. Long waiting time in any hospital is considered as an indicator of poor quality and needs improvement. Managing waiting lines create a great dilemma for managers seeking to improve the return on investment of their operations. Customers also dislike waiting for long time.

The study is designed to help the management of Essikado Hospital on the employee adequacy and also help to reduce patient-waiting time for services at the dental department. Hence, this introductory chapter of the research involves: the Background of the study, Problem statement, Objectives of the study, Research methodology, Justification of the Study, and the Organization of the study.

1.2 Background of the Study

A queue is a waiting line, whether of people, signals or things (Ashley, 2000). Queuing time is the amount of time a person, a signal or a thing spends before being attended to for services. Queuing theory is generally considered as a branch of operations research because its results are often used when making business decisions about the resources needed to provide services. The theory seeks to determine how best to design and operate a system usually under conditions requiring allocation of scarce resources. Queuing models are those where a facility performs a service.

A queuing problem arises when the current service rate of a facility falls short of the current flow rate of customers. If the size of the queue happens to be a large one, then at times it discourages customers who may leave the queue and if that happens, then a sale is lost by the concerned business unit. Hence, the queuing theory is concerned with the decision making process of the business unit which confronts with queue questions and makes decisions relative to the numbers of service facilities which are operating. The earliest use of queuing theory was in the design of a telephone system. A queue can be studied in terms of the source of each queued item, how frequently items arrive on the queue, how long the item can or should wait, whether some items should jump ahead in the queue, how multiple queues might be formed and managed, and rules by which items are queued or de-queued.

The queuing theory enables mathematical analysis of several related processes, including arriving at the (back of the) queue, waiting in the queue (essentially a storage process), and being served at the front of the queue. The queuing theory permits the derivation and calculation of several performance measures including the service, and the probability of encountering the system in certain states such as empty, full, having an available server or having to wait a certain time to be served (Vohra, 2010).The subject of queuing theory can be described as follows: consider a service centre and population of customers, which at sometimes enter the system in order to obtain service. It is often the case that the service centre can only serve a limited number of customers. If a new customer arrives and the service is exhausted, he/she enters a waiting line and waits until the service facility becomes available. So we can identify three main elements of service centre: a population of customers, the service facility and the waiting line.

Queues are formed because of limited resources and they are experienced almost everyday. Thus queuing theory calculates the average time a customer spends in a system, the average time a customer spends in a line, and the average time a customer spends in service. Queuing theory is applicable to intelligent transportation systems, call centres and advanced telecommunications. Queuing Theory tries to answer questions like e.g. the mean waiting time in the queue, the mean system response time (waiting time in the queue plus service times), mean utilization of the service facility, distribution of the number of customers in the queue, distribution of the number of customers in the system and so forth. These questions are mainly investigated in a stochastic scenario, where e.g. the inter arrival times of the customers or the service times are assumed to be random.

Areas where queuing theory could be applied in healthcare centres include:

1. Public health Queuing models can be used for public health. For example, the resources needed for Mass Vaccination camp in a particular area, facility

and resource planning for emerging or changing disease profiles or changing demographics.

2. Health care resource and infrastructure planning for disaster management Any type of disaster cause significant human and economic damage and they all demand a crisis response. It demands immediate rescue of people, provision of medical services needed and containment of the damage to people and property. In such scenarios, queuing models are frequently used in conjunction with simulation to answer the what-if questions, to plan, organize and be prepared for the calamities.

3. Hospital pharmacy and pharmacy stores In pharmacy, queuing theory can be used to assess a multitude of factors such as prescription time, patient waiting time, patient counselling- time and staffing levels. The application of queuing theory may be a particular benefit in pharmacies with high volume outpatient workloads and or those that provide multiple points of service. By better understanding queuing theory, service managers can make decisions that increase the satisfaction of all relevant group-customers, employees and management.

4. Walk-in-patients in physician offices, out-patient clinics and out-patient surgeries in hospitals.

The most common objectives of studies on the clinics have included the reduction of patients time in the system, improvement on customer service, better resource utilization and reduction of operating costs. (Bharali,2010).

1.3 Dental Clinic

Dentistry is a healthcare specialism that is often forgotten about or given less importance than other areas. However as diets change to include more sugary snacks it has fewer fruits and vegetables due mainly to globalisation and a wider variety of goods, there is an increasing need for people to have regular dental check-ups and sometimes even actual dental treatment. Unfortunately Ghana, like most developing countries, is not blessed with numerous supply of dentist. There are only a few dentists in Ghana to meet the needs of the Ghanaian population especially in the rural areas.

A dental clinic is an integral part of a social and medical organization, the function of which is to provide for the population dental care and whose outpatient services reach out to the family and its home environment. Nowadays, there are a lot of dental clinics in the Sekondi-Takoradi Metropolis. This is due to the fact that a lot of the dentist in the government hospitals in Sekondi-Takoradi have opened their private dental clinics thereby reducing the pressure in the hospitals. Some of the services provided include pain relief (tooth extraction, temporary incision and drainage) and dental restoration (simple amalgam filling, temporary dressing).

1.4 A Queuing System

Examples of queuing abound in daily life: queuing situations at a ticket window in the railway station or post office, at the cash points in the supermarket, the waiting room at the airport, train or hospital, etc. In telecommunications, the packets arriving at the input port of a router or switch are buffered in the output queue before transmission to the next hop towards the destination. In general, a queuing system consists of (i) arriving items (packets or customers), (ii) a buffer or waiting room, (iii) a service center and (iv) departures from the system. The main processes are stochastic in nature. Initially in queuing theory, the main stochastic processes were described in continuous time, while with the introduction of the Asynchronous Transfer Mode (ATM) at the late eighties; many queuing problems were more effectively treated in discrete time, where the basic time unit or time slot was the minimum service time of one ATM cell. In the literature, there is unfortunately no widely adopted standard notation for the main random variables, which often troubles the transparency.

1.5 Basics of Queuing Theory

For application of queuing models to any situation we should first describe the input and output processes. The table below gives some examples of input and output processes.

Setting	Input	Output
Hospital	Arrival of patients	Assessment, triage, provision of services, discharge
Bank	Arrival of Customers	Provision of services by tellers
Supermarket	Shoppers	Checkout centers

1.6 Characterization

A queuing system may be described as a system, where customers arrive according to an arrival process to be serviced by a service facility according to a service process. Each service facility may contain one or more servers. It is generally assumed that each server can only service one customer at a time. If all servers are busy, the customer has to queue for service. If a server becomes free again, the next customer is picked from the queue according to the rules given by the queuing discipline. During service, the customer might run through one or more stages of service, before departing from the system. In queuing theory, models are used to describe the characteristics of a queuing system. Some of the more commonly considered characteristics are discussed below.

(i). **Queue Length**: : The average number of customers in the queue waiting to get service. Large queues may indicate poor server performance while small queues may imply too much server capacity.

(ii). **System Length**: The average number of customers in the queue in the system, those waiting to be and those being serviced. Large values of this statistic imply congestion and possible customer dissatisfaction and potential need for greater service capacity.

(iii). **Waiting time in the queue**: The average time that a customer has to wait in the queue to get service. Long waiting times are directly related to customer dissatisfaction and potential loss of future revenues, while very small waiting times may indicate too much service capacity.

(iv). **Total time in the system**: The average time that a customer spends in the system, from entry in the queue to completion of service. Large values of this statistic are indicative of the need to make adjustment in the capacity.

(v). **Service idle time**: The relative frequency with which the service system is idle. Idle time is directly related to cost. However, reducing idle time may have adverse effects on the other characteristics mentioned.

It is important to mention here that the results obtained from various models are based on the assumption that the service the service system is operating under equilibrium or steady state conditions. For many systems, the operating day begins in transient state with no customers in the system. It takes some initial time interval for enough customers to arrive such that a steady state does not mean that the system will reach a point where the number of customers in the system never changes. Even when the system reaches equilibrium, fluctuations will occur. A steady state condition really implies that various system performance measures (the operating characteristics) would reach stable values.

A queuing system is characterized by four components:

(i). Arrival process

(ii). Output process

(iii). Service mechanism and

(iv). Queue discipline. (Vohra, 2010).

1.6.1 Arrival process (Input process)

Arrivals may originate from one or several sources referred to as the calling population. The calling population can be limited or 'unlimited'. An example of a limited calling population may be that of a fixed number of machines that fail randomly. The arrival process consists of describing how customers arrive to the system. Arrivals from the calling population may be classified on different bases as follows:

(i). **According to source** - the source of customers for a queuing system can be infinite or finite. For example, all people of a city could be potential customers at a supermarket. The number of people being very large, it could be taken to be infinite. On the other hand, there are many situations in business and industrial conditions where we cannot consider the population to be infinite.

(ii). **According to numbers** - the customers may arrive for service individually or in groups. Single arrivals are illustrated by customers visiting a beautician. On the other hand, families visiting restaurants, ship discharging cargo at a dock are examples of bulk, or batch, arrivals.

(ii). **According to time** - customers may arrive in the system at known (regular or otherwise) times, or they might arrive in a random way. The queuing models wherein customers' arrival times are known with certainty are categorized as deterministic models and are easier to handle. On the other hand, a substantial majority of the queuing models are based on the premise that the customers enter the system stochastically, at random points in time.

With random arrivals, the number of customers reaching the system per unit time might be described by a probability distribution. Although the arrivals might follow any pattern, the frequently employed assumption, which adequately supports many real world situations, is that the arrivals are Poisson distributed. (Vohra, 2010).

1.6.2 Output process

To describe the output process we usually specify the service time distribution (service mechanism). The service mechanism is the way customers receive service once they are selected from the queue. The output process tells the time that a customer leaves the system after going through all the service mechanisms.

1.6.3 Service mechanism

The service mechanism of a queuing system is specified by the number of servers (denoted by s), each server having its own queue or a common queue and the probability distribution of customer's service time. Basically there are about four types of service mechanism and different combinations of the same can be used for very complex networks. They are:

(i) **A single service facility**: in this system there is only one queue and only one server. Here customers in the queue wait till the service point is ready to take them for servicing.

(ii) **Multiple, parallel facilities with single queue**: That is, there is more than one server. The term parallel implies that each server provides the same type of facility. Example, booking at a service station that has several mechanics, each handling one vehicle at a time.

(iii) **Multiple, parallel facilities with multiple queue**: This model is different from the multiple, parallel facilities with single queue in that each of the servers has different queue. Example is, different cash counters in an account office where students can make payment of their school fees.

(iv) **Service facilities in a series**: In this, a customer enters the first station and gets a portion of service and then moves on to the next station, gets some service and then again moves on to the next station and so on, and finally leaves the system, having received the complete service. For example, machining of a certain steel item may consist of cutting, turning, knurling, drilling, grinding and packaging operations, each of which is performed by a single server in a series.

1.6.4 Queue Discipline

Discipline of a queuing system means the rule that a server uses to choose the next customer from the queue (if any) when the server completes the service of the current customer. The most common used queue disciplines are:

(i) **First-Come-First-Serve (FCFS)**: Customers are served on first-come-first-served basis. This is the most common discipline used in queuing systems. For example, with a queue at the bus stop, the people who came first will have their tickets first and board the bus first.

(ii) **First-In-First-Out (FIFO)**: Customers are served on a first-in first-out basis. For example, with a queue at the hospital, a patient who comes in first is served and leave before the others.

(iii) **Last-Come-First-Serve (LCFS)**: That is customers who come in last are served first. Here, the customers are serviced in an order reverse of the order in which they enter so that the ones who join the last are served first. For example, the people who join an elevator last are the first ones to leave it. Also, assume that letters to be typed, or order forms to be processed accumulate in a pile, each new addition being put on the top of them. Thus, a just arriving task would be the next to be serviced provided that no fresh task arrives before it is picked up.

(iv) **Priority**: The customers in a queue might be rendered service on a priority basis. Thus, customers may be called according to some identifiable characteristics for service. Example is treatment given to a very important personality in preference to other patients in a hospital.

McQuarrie (1983) showed that it is possible, when utilization is high, to minimize waiting times by giving priority to clients who require shorter service times. This rule is a form of the shortest processing time rule that is known to minimize waiting times. It is found infrequently in practice due to the perceived unfairness (unless that class of customers is given a dedicated server, as in supermarket check-out systems) and the difficulty of estimating service times accurately.

The System Capacity (Size of the Source Population)

The system capacity is the maximum number of customers, both those in service and those in the queue(s), permitted in the service facility at some time. Whenever a customer arrives at a facility that is full, the arriving customer is denied entrance to the facility. Such a customer is not allowed to wait outside the facility (since that effectively increases the capacity) but is forced to leave without receiving service.

A system that has no limit on the number of customers permitted inside the facility has infinite capacity. Examples of infinite capacities include shoppers arriving at a supermarket, cars arriving at a highway toll booth, students arriving to register courses at a large University. Most queuing models assume such infinite arrival population. A system with a limit on the number of customers has finite capacity.

1.7 Attitude of Customers In the Queuing System

Customers in the queuing system can be classified as being patient or impatient. If a customer joins a queue if it exists, and wait till they enter the service station for getting service, they are called patient customers. On the other hand, the queuing systems enjoy customer behaviour in the form of defections from the queue. There may be jockeying among the many queues that is a customer may switch to the queues which are moving fast. Reneging is also possible in the queue. Here a customer stands in the queue for some time and then leaves the system because it is working slowly. The probability that a patient reneges usually increases with the queue length and the patient's estimate of how long he must wait to be served. In systems where demand exceeds server capacity, reneging is the only way that a system attains a 'state of dysfunctional equilibrium' (Hall et al., 2006).

Some customers on the other hand, may decide not to join the queue for some reason and may decide to return for the service later and this situation is known as balking. Blocking occurs when a queuing system places a limit on queue length. For example, an outpatient clinic may turn away walk-in patients when its waiting room is full. In a hospital, where in-patients can wait only in a bed, the limited number of beds may prevent a unit from accepting patients.

McManus et al., (2004) presented a medical-surgical Intensive Care Unit where critically ill patients cannot be put in a queue and must be turned away when the facility is fully occupied. This is a special case where the queue length cannot be greater than zero, which is called a pure loss model (Green, 2006a).

Koizumi et al., (2005) found that blocking in a chain of extended care, residential and assisted housing facilities results in upstream facilities holding patients longer than necessary. They analysed the effect of the capacity in downstream facilities on the queue lengths and waiting times of patients waiting to enter upstream facilities. System-wide congestion could be caused by bottlenecks at only one downstream facility.

1.8 Statement of the Problem

The Essikado Hospital is one of the hospitals in the Sekondi-Takoradi metropolis that has a dental department. The hospital dental department was started in 1991. The dental department offers a full range of all dental health services which include all types of fillings (amalgams, composite, and Root cal treatment), extractions, orthodontic treatment, and prosthetic treatment. The dental department seeks to:

(i). Provide quality healthcare

(ii). Increase access to dental service

(iii). To improve upon its dental service delivery

(iv). To reduce patients waiting time.

The challenge at the dental department is that the queue is long and service is very slow.

1.9 Objectives

The main purpose of this study is to look at the dental department of the Essikado Hospital. Specifically, this research work seeks to find:

(a). The average number of arrivals entering the dental department of the hospital

(b). The average service time of customers at the dental department.

(c). The average time a patient spends in the queue.

hospital.

1.10 Methodology

The data used for this study were obtained from the Essikado Hospital. The arrival time, the time a patient is called to see the dentist and the time service ended for patients were recorded. The data were gathered for a period of two weeks (13th-24th April 2015) The queuing discipline used for this study was the first-come-first served discipline. Sources of reading material were the internet and the K.N.U.S.T library. The method of solution is manual. The queuing model used was the single server queuing model.

1.11 Significance of the Study

The findings of this study will assist the hospital administrators and management to improve on customer service, and maximize the utilization of its resources (doctors, nurses, hospital beds, etc.). The findings could also be used for appropriate staffing and facilities design.

1.12 Organization of the Study

The study is organized into five chapters. Chapter one deals with the background, problem statement, objectives, scope, limitations and research methodology of the study. Chapter two presents the review of pertinent literature. In chapter three, we shall put forward the research methodology of the study. Chapter four is devoted for data collection and analysis. Chapter five which is the final chapter of the study presents summary of findings, conclusions and recommendation.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

This chapter reviews relevant and adequate literature on queuing theory. It contains other research findings that will assist the researcher and serve as a guide for future research. The chapter also highlights on the history behind the theory being used the basics of queuing theory.

2.2 Brief History of Queuing Theory

Queuing theory was developed to provide mathematical models to predict behaviour of systems that attempt to provide service for randomly arising demands and can trace its origins back to a pioneer investigator, Danish mathematician named A. K. Erlang, who, in 1909, published The Theory of Probabilities and Telephone Conversations based on work he did for the Danish Telephone Company in Copenhagen, Denmark. Work continued in the area of telephone applications, and although the early work in queuing theory picked up momentum rather slowly, the trend began to change in the 1950s when the pace quickened and the application areas broadened well beyond telephone systems.

In Erlang's work (1990), as well as the work done by others in the twenties and thirties, the motivation has been the practical problem of congestion. An example is the works of Molina(1927). During the next two decades other researchers became interested in these problems and developed general models which could be used in more complex situations. Noting the inadequacy of the equilibrium theory in many queue situation, Pollaczek, 1965, began investigations of the behaviour of the system during a finite time interval. Since then and throughout his career, he did considerable work in the analytical behavioural study of queueing systems (Pollaczek, 1965). The trend towards the analytical study of the basic stochastic processes of the system continued, and queuing theory proved to be a fertile field for researchers who wanted to do fundamental research on stochastic processes involving mathematical models.

Most of the pioneers of queuing theory were engineers seeking solutions to practical real world problems. The worth of queuing analysis was judged on model usefulness in solving problems rather than on the theoretical elegance of the proofs used to establish their logical consistency. To this day a large majority of queuing theory results used in practice are those derived under the assumption of statistical equilibrium.

2.3 Psychology of Queuing

The mathematics of queuing theory enables a decision- maker to model the behaviour of a queuing system. Mathematical equations are used to calculate the time spent waiting and the number of customers waiting. In some situations, if the lines are too long, customers may go elsewhere to be served. In this case the queuing manager is interested in the number of customers lost due to long waits. Generally, if the waiting time seems excessive and customers are dissatisfied, a manager will explore cost-effective strategies for increasing the service capacity by adding more servers or increasing the speed of service. However, raw numbers fail to tell the whole story. The experience of waiting in line is influenced by the waiting area environment and our expectations as to the length of the wait. Imagine having to wait standing up in a dentist's office for twenty minutes, while a patient is screaming in an adjacent examination room. Now imagine an alternative wait in comfortable chairs with access to the 'latest' magazines for a variety of customer tastes.

Many companies (Disney is one example) have become expert in understanding the psychology of waiting. Waiting in a line that is moving seems less boring than standing still in the same spot. TV monitors with engaging pictures help keep visitors' minds off the clock the clock. In addition, if they can see and hear some of the excitement of those who have completed their wait, anticipation increases and waiting seems worthwhile. Lastly, expectations are a major factor in determining customer satisfaction. If customers approach a line and are told that the wait will be fifteen minute, at least they have the information to make an informed judgement as to join the line or not. If it turns out to be less than the quoted fifteen minutes then they are pleasantly surprised. Another dimension to the psychology of waiting relates to fairness. It can be very upsetting to see someone arrive after you in line and end up being served before you. This can happen if there are two separate lines. You might get stuck behind a customer who has a complicated request that takes a long time to service. As a result, people who have joined the other line even after you might end up waiting less time. Many organizations have addressed this potential inequity by creating one line which all arriving customers enter. Thu, anyone who arrives after you must be further back in line and cannot begin service before you do.

2.4 Related Works On Queuing Theory

Yok et al (2005) indicated that patient satisfaction in a dental clinic is concerned with meeting clients perceived needs and concerns. The dental patients' needs and concerns not only include considerations for the technical quality of the service, convenience of the service, friendly atmosphere, cleanliness of the environment and equipment but also includes respect for the time patients spend in the clinic if it is to retain its clients as well as attract others.

Gorunescu et al (2002) designed a queuing model for bed-occupancy management and planning of hospitals. The aim of the paper was to present the movements of patients through a hospital department by using classical queuing theory and, on the other hand, presented a way of optimising the use of hospital resources in order to improve hospital care. The methodology developed enabled estimation of main characteristics of access to service for patients and hospital managers: the probability of lost demands and the mean number of patients in the hospital department (bed occupancy). The paper determined optimal bed numbers for a hospital system where the writers describe patient arrivals by a Poisson process, hospital beds as the servers and length of stay were modelled using phase-type distributions.

Agnihothri and Taylor (1991) sought the optimal staffing a hospital scheduling department that handles phone calls whose intensity varies throughout the day. There are known peak and non-peak periods of the day. The authors grouped the periods that receive similar call intensity and determines the necessary staffing varies dynamically with call intensity. As a result of redistributing server capacity over time, customer complaints immediately reduce without an addition of staff.

Ajay et al (2013) on the other hand stressed the usefulness of queuing systems in the society and that the capacity of these systems can have an important result on the quality of human life and productivity of the process. They described the scheduling discipline and queuing networks. Their analysis provides fundamental information for successfully designing systems that achieve an appropriate balance between the cost of providing a service and the cost associated with waiting for that service.

Dhar and Rahman (2013) worked on a case study for Bank ATM Model. The authors used queuing theory to study the waiting lines in Brac Bank ATM in Bangladesh. Little's theorem and single server queuing model were used to derive the arrival rate, service rate, utilization rate and waiting time in the queue. They concluded that the rate at which customers arrive in the queue system is 1 customer per minute and the probability of buffer overflow is the probability that customers will run away, because maybe they are impatient to wait in the queue.

Nuti et al(2012) also researched on Managing waiting times in diagnostic medical imaging. The paper analysed the variation in the delivery of diagnostic imaging services in order to suggest possible solutions for the reduction of waiting times, increase the quality of services and reduce financial cost. Waiting times measured per day were compared on the basis of the variability in the use rates of CT and MRI examinations as well as of the number of radiologists available.

Hall et al (2006) classified the queue problems into three types:

1. The Perpetual Queue: It is the worst type of queue. It is a queue for which all customers have to wait for service. All customers have to wait because the servers have insufficient capacity to handle the demand for the service (e.g. government service).

2. The Predictable Queue: This occurs when the arrival rate is known to exceed the service capacity over finite interval of time (e.g. the rush-hour traffic jam and lunchtime at restaurants).

3. The Stochastic Queue: It is not predictable. They occur by chance when customers happen to arrive at a faster rate than they are served. Stochastic queues can occur whether or not the service capacity exceeds the average arrival rate.

Mackey and Cole, 1997 indicated that patient waiting time is generally known as the length of period taken from when the patient enters the waiting room or the consulting room until when the patient actually leaves the hospital/clinic. Due to the fact that dental procedures take varying time depending on the complexity of the treatment procedure, in their study, patient waiting time specifically referred to the length of period from registration of the patient to the time when he/she was called into the surgery unit to start treatment.

Ndukwe et al (2011) used queuing models to work on reducing queues in a Nigerian hospital pharmacy. The aim of the work was to characterize the queue, describe the queue discipline of the outpatient pharmacy, to institute a cross-sectional intervention by streamlining queue behaviour and to measure the impact of streamlining queue characteristics and queue discipline on waiting time of patients. When the intervention was done involving staff re-orientation, the streamlined process reduced waiting time. Queue discipline was strictly instituted by designed tally cards that were serially numbered. The authors concluded that effort should be intensified by hospital pharmacists to reduce patient queues and improve efficiency of services.

Kembe et al (2012) used queuing theory to study waiting and service costs of a multi-server queuing model in a specialist hospital. The waiting and service costs were determined with a view to determining the optimal service level. Data were collected through observations, interview and by administering questionnaires. TORA optimization software was used to analyse the results. The authors indicated from the analysis that the average queue length, waiting time of patients as well as over-utilization of doctors could be reduced when the service capacity level of doctors at the clinic is increased.

Muhammad et al (2014) researched on measuring queuing system and time standards: a case study of student affairs in universities. The objective was to examine the behaviour and patterns of arrival of students in a university through observation. The authors adopted queuing theory for waiting in lines/queues. They observed that students usually wait for minutes, hours, days or months to receive desired service for which they are waiting. They indicated that there are two ways to increase customers' satisfaction with regard to waiting time: by decreasing actual waiting time, and through enhancing customer's waiting experience. If the organizations cannot control the actual duration of the waiting, then it might consider how it manipulate the perceived wait time

Siddhartan et al (1996) analysed the effect on patient waiting times when primary care patients use the Emergency Department. The authors proposed a priority discipline for different categories of patients and then a first-in-first-out discipline for each category. The authors found that the priority discipline reduces the average wait time for all patients, however, while the wait time for higher priority patients reduces, lower priority patients endure a longer average waiting time. The amount of time that the patient spends at the health facility has often been used as a measure of the patients' satisfaction with the service being provided.

Anokye et al (2013) seeks to find a basic model of vehicular traffic based on queuing theory. The authors used the infinite server queuing model to model the vehicular traffic flow and to explore how vehicular traffic could be minimized using queuing theory in order to reduce delays on roads in Kumasi. The queue discipline use first-come-first- served. Computations of queuing parameters were done manually and the system was assumed to reach a steady state, a condition that the rate of arrival and service is constant.

Yariv et al (2012) in their paper, designing patient flow in emergency departments, developed a methodology for Emergency Department (ED) design. Data were collected over a period of two to four years from eight hospitals, of various sizes and deploying various ED operational models. A simulation model developed by Sinreich and Marmor (2005), which already validated their model on the relevant hospitals was used to extend the scope of analysis. They discovered that different operational models have weaknesses and strengths over various uncontrollable parameters.

Davis and Vollman (1990) stressed that patient evaluation of service quality is affected not only by the actual waiting time but also by the perceived waiting time. The act of waiting has significant impact on patients satisfaction. The authors concluded that the amount of time customers must spend waiting can significantly influence their satisfaction.

Shimshak et al (1981) considered a pharmacy queuing system with preemptive service of prescription order that suspends the processing of lower priority prescriptions. Different costs are assigned to wait-times for prescriptions of different priorities.

Queuing theory uses mathematical models and performance measures to assess and hopefully improve the flow of customers through a queuing system (Bunday, 1996). A good patient flow means that the patient queuing is minimized while a poor patient flow means patients suffer considerable queuing delays. McClain (1976) reviews research on models for evaluating the impact of bed assignment policies on utilization, waiting time and the probability of turning away patients.

Nosek and Wilson (2001) review the use of queuing theory in pharmacy applications with particular attention to improving customer satisfaction. Customer satisfaction is improved by predicting and reducing waiting times and adjusting staffing.

Preater (2002) presents a brief history of the use of queuing theory in healthcare and points to an extensive bibliography of the research that lists many papers (however, it provides no description of the applications or results). Green (2006) presents the theory of queuing as applied in healthcare. She discusses the relationship amongst delays, utilization and the applications of the theory to determine the required number of servers.

Bank et al (2001), indicted that delays and queuing problems are most common features not only in our daily life situations such as at a bank or postal office, at a ticketing office, in public transportation or in a traffic jam but also in more technical environments, such as in manufacturing, computer networking and telecommunications. They play an essential role for business process re-engineering purposes in administrative tasks. Queuing models provide the analyst with a powerful tool for designing and evaluating the performance of queuing systems. The authors also observed that whenever customers arrive at a service facility, some of them have to wait before they receive the desired service. It means that the customer has to wait for his/her turn, may be in a line. Customers arrive at a service facility (sales checkout zone in ICA) with several queues, each with one server (sales checkout counter). The customers choose a queue of a server according to some mechanism (e.g., shortest queue or shortest workload).

Adan et al (2000) iterated that sometimes, insufficiencies in services also occur due to an undue wait in service may be because of new employee. Delays in service jobs beyond their due time may result in losing future business opportunities. Queuing theory is the study of waiting in all these various situations. It uses queuing models to represent the various types of queuing systems that arise in practice. The models enable an appropriate balance between the cost of service and the amount of waiting. Managing waiting lines create a great dilemma for managers seeking to improve the return on investment of their operations. Customers also dislike waiting for long time. If the waiting time and service time is high customers may leave the queue prematurely and this in turn results in customer dissatisfaction. This will reduce customer demand and eventually revenue and profit.

Biju et al (2011)indicated that though queuing analysis has been used in hospitals and other healthcare settings, its use in this sector is not widespread. With rapid change and realignment of healthcare system, new lines of services and facilities to render the same, severe financial pressure on the healthcare organizations, and extensive use of expanded managerial skills in healthcare setting, use of queuing models has become quite prevalent in it. In an era of healthcare reform, improving quality and safety, and decreasing healthcare cost have become even more important goals than before. With rapid change and realignment of healthcare system, new lines of services and facilities to render the same, severe financial pressure on the healthcare organizations, and extensive use of expanded managerial skills in healthcare setting. Scientific management of patient flow is at the heart of our ability to achieve these goals. While on one hand we are faced with overcrowded facilities, on the other hand, the industry's financial conditions do not allow us to add resources liberally. At its most basic level, queuing theory involves arrivals at a facility (i.e. computer store, pharmacy, bank) and service requirements of that facility (i.e., technicians, pharmacists, tellers). The number of arrivals generally fluctuates over the course of the hours that the facility is available for business. The theory enables mathematical analysis of several related processes, including arriving at the (back of the) queue, waiting in the queue (essentially a storage process), and being served at the front of the queue. The theory permits the derivation and calculation of several performance measures including the average waiting time in the queue or the system, the expected number waiting or receiving service, and the probability of encountering the system in certain states, such as empty, full, having an available server or having to wait a certain time to be served.

Queuing management consists of three major components:

1. How customers arrive

2. How customers are serviced

3. The condition of the customer exiting the system

Arrivals: Arrivals are divided into two types:

1. Constant - exactly the same time period between successive arrivals (i.e., machine controlled).

2. Variable - random arrival distributions, which is a much more common form of arrival.

A good rule of thumb to remember the two distributions is that time between arrivals is exponentially distributed and the numbers of arrivals per unit of time is Poisson distributed.

Hillier and Lieberman, (1974) were of the view that servicing or queuing system consists of the line(s) and the available number of servers. Factors to consider include the line length, number of lines and the queue discipline. Queue discipline is the priority rule, or rules, for determining the order of service to customers in a waiting line. The authors indicated that one of the most common used priority rules is first-come-first served, (FCFS). Others include a reservations first, treatment via triage (i.e., emergency rooms of health centres), highest-profit customer first, largest orders first, best customers first and longest wait-time first. An important feature of the waiting structure is the time the customer spends with the server once the service has started. This is referred to as the service rate: the capacity of the server in numbers of units per time period (i.e., 15 orders per hour).

Winston and Albright (1997) states that the usual queue discipline is first come, first served (FCFS or FIFO), where customers are served in order of arrival. In this study the dental uses FCFS queuing discipline. Although, sometimes there are other service disciplines: last come, first served (which happens sometime in case of emergencies), or service-in-random order and priority rule.

Davis and Acquilano (2003) assert that reservations first, emergencies first, highest profit customer first, largest orders first, best customers first, longest waiting time in line, and soonest promised date are other examples of queue discipline. Unless otherwise stated, the queuing model adopted in this study assumes arrival from infinite source with infinite queue and with first in first served (FCFS) queue discipline.

According to Nosek and Wilson (2001), service mechanism describes how the customer is served. In a single server system each customer is served by exactly one server, even though there may be multiple servers. In most cases, service times are random and they may vary greatly. Sometimes the service time may be similar for each job or constant. The service mechanism also describes the number of servers. A queuing system may operate with a single server or a number of parallel servers. An arrival who finds more than one free server may choose at random any one of them for receiving service. If he finds all the servers busy, he joins a queue common to all servers. The first customer from the common queue goes to the server who becomes free first.

According to Medhi (2003) a system may have an infinite capacity, that is, the queue in front of the server(s) may grow to any length. Furthermore, there may be limitation of space and so when the space is filled to capacity, an arrival will not be able to join the system and will be lost to the system. The system is called a delay system or a loss system, according to whether the capacity is infinite or finite respectively. Once customers are served, they depart and may not likely re-enter the system to queue again. It is usually assumed that departing customers do not return into the system immediately .

Chase et al.,(2004) is of the opinion that once a customer is served, two exit fates are possible:

1. The customer may return to the source population and immediately become a competing candidate for service again.
2. There may be a low probability of re-service. In hospitals, departure means home discharge.

In addition, patient flow times play an increasingly important role in today's healthcare systems. Government reimbursement systems (based on a justified length of stay), insurer's rejection of reimbursement (i.e. denied days), competition between hospitals, government regulations and patient satisfaction urge hospital decision makers to find ways to decrease waiting times (both waiting time inside the hospital as well as the waiting list that exists outside the hospital). Current healthcare literature and practice indicate that waiting lists and congested patient flows do indeed make up for one of the most important problems in care industries.

Adedayo et al (2006) indicated that in order to improve performance in an environment as complex as a hospital system, the dynamics at work need to be understood. To obtain such an understanding, queuing theory and simulation provide an ideal set of tools. Healthcare systems however have a number of specific features as compared to manufacturing systems, posing important methodological challenges. In healthcare, queuing theory has been used to assess capacity requirements , outpatient scheduling, process optimization and absence recovery modelling when the capacity of service provided fall short of the demand for the service.

Sanish (2007) in his article on application of queuing to the traffic at New Mangalore Port refers to queuing theory as an analytical techniques accepted as valuable tool for solving congestion problems. According to him the primary inputs to the models are the arrival and service patterns. These patterns are generally described by suitable random distribution. He observed that the arrival rate of ships follows exponential distribution while the service time follows Erlang or Poisson distribution. He observed that queuing theory can be used to predict some important parameters like average waiting time of ships, average queuing length, average number of ships in the port and average berth utilization factor closer to the actual values.

Stakutis and Boyle,(2009) proposed that the effect of queuing during hospital visits in relation to the time spent for patients to access treatment in hospitals is increasingly becoming a major source of concern to a modern society that is currently exposed to great strides in technological advancement and speed. Internal operational factors majorly determine outpatient waiting times and include; arrival pattern of prescriptions, sequencing of work, percentage of staff at work, interaction between the pharmacy service providers that is assessor and technician interaction or technician and counsellor interaction.

In many factories, queue time constitutes about 90% of the total lead time (Crabtree, 2008). Queue time can be extended to hospitals pharmacies where patients complain about the amount of time spent before treatment can be received. Queues deal with problems which involve waiting for a product or service. Typical examples might be hospital/pharmacies/banks/supermarkets waiting for service, computers used in Hospital Management Information System (HMIS) waiting for response, power failure situation that involves waiting for electric power to be restored to enable a piece of machinery used for compounding a prescription medication function. Another example is the public transport system- waiting for a train or a bus to convey a consultant pharmacist to the hospital in the advent of an accident, poisoning or hypertensive emergency.

Woensel and Cruz (2009) indicated that queuing models have mainly been used to study congestion for interrupted traffic flows at signalized and un-signalized intersections. However, they also proposed that it can also be usefully applied to describe and analyse congestion for uninterrupted traffic flows in spite of the absence of the development of a formal queue. Queuing models applied to traffic situations provide an adequate description of the complex dynamic and stochastic environment under study. The authors suggested that in designing queuing systems we need to aim for a balance in services rendered to clients. In essence, all queuing systems can be broken down into individual subsystems, consisting of entities queuing for some activity e.g. dispensing and counselling.

David (2005) suggested how queue characteristics determines how, from the set of clients waiting for service, we choose the one to be served next (e.g. FIFO (first-in-first-out); LIFO (last-in-first-out); randomly). This order instituted for waiting is often called the queue discipline. Queue discipline can be examined so as to determine the queue characteristics to implore over a given period in providing pharmaceutical services from the outpatient section or from any other outlet. The authors indicated that queue discipline can include balking (clients or patients deciding not to join the queue if it is too long), reneging (clients or patients leave the queue if they have waited too long for service), jockeying (clients or patients switch between queues if they think they will get served faster by so doing).

2.5 Attitude of Customers in the Queuing system

Hall et al. (2006) classified customers in the queuing system as being patient or impatient. If a customer joins a queue if it exists, and wait till they enter the service station for getting service, they are called patient customers. On the other hand, the queuing systems enjoy customer behaviour in the form of defections from the queue. There may be jockeying among the many queues that is a customer may switch to the queues which are moving fast. Reneging is also possible in the queue. Here a customer stands in the queue for some time and then leaves the system because it is working slowly. The probability that a patient reneges usually increases with the queue length and the patient's estimate of how long he must wait to be served. In systems where demand exceeds server capacity, reneging is the only way that a system attains a 'state of dysfunctional equilibrium'. Some customers on the other hand, may decide not to join the queue for some reason and may decide to return for the service later and this situation is known as balking

Blocking occurs when a queuing system places a limit on queue length. For example, an outpatient clinic may turn away walk-in patients when its waiting room is full. In a hospital, where in-patients can wait only in a bed, the limited number of beds may prevent a unit from accepting patients.

McManus et al., (2004) presented a medical-surgical Intensive Care Unit where critically ill patients cannot be put in a queue and must be turned away when the facility is fully occupied. This is a special case where the queue length cannot be greater than zero, which is called a pure loss model

Koizumi et al., (2005) found that blocking in a chain of extended care, residential and assisted housing facilities results in upstream facilities holding patients longer than necessary. They analysed the effect of the capacity in down-stream facilities on the queue lengths and waiting times of patients waiting to enter upstream facilities. System-wide congestion could be caused by bottlenecks at only one downstream facility

As analysed by Fink and Gillett (2006) the cost of a dissatisfied customer is not negligible, they described Waiting in line is a primary source of dissatisfaction. They mentioned that a well-known queuing theory and integrating theory behind the Taguchi Loss Function, a manager can derive the costs associated with this dissatisfaction and that customer dissatisfaction is not just an issue at the upper specification limit, but rather for each moment in time beyond the targeted wait time. They illustrated by using the Taguchi Function, it can then be seen that these costs increase beyond the upper specification limit. However, by assessing these costs and then taking measures to reduce either the actual or perceived waiting times, organizations can quantitatively determine the cost-benefit relationship of improved waiting lines.

Chin (2007), investigated the submittal review/approval process and used queuing theory to determine the major causes of long lead times. Under his study, he explored the underlying causes of waiting in a process flow and found the improvement methods from the queuing perspective.

Yeung et al (2002), conducted research on large hospitals with the help of Laplace transform of the probability density function of customer response time in networks of queues with class-based priorities. He obtained the mean & standard deviation of total patient service time for large hospitals mainly for accident and emergency department.

Georgiveskiy et al (2002) examined the widespread context problem of extended waiting times for health services in the context of the Emergency Department (ED) at a regional hospital. The authors also examined Operations Research (OR) to reduce the waiting time in the hospital-admitting department. They conducted their study in five faces, from the collection of data to actual improvement in the quality of the health care delivery system.

According to Koizumi (2002) the current trend toward downsizing and closing of state mental health institutions has led to an over-utilization of many local mental health facilities. This problem has often been intensifying by a shortage of long-stay psychiatric hospitals and community-type accommodations. He specifically examines blocking in interrelated mental health facilities, including

(a) acute hospitals (where patients wait to enter extended acute hospitals),

(b) extended acute hospitals (where patients wait to enter residential facilities), and

(c) residential facilities (where patient wait to enter supported housing).

Singh (2006) found that the queuing theory in healthcare organizations is very beneficial. He used Queuing model to achieve a balance or trade-off between capacity and services delays and used the POM-QM Software to demonstrate it. Fomundam and Herrmann (2007) described the contributions and applications of queuing theory in the field of healthcare. They summarized a range of queuing theory results in areas of waiting time and utilization analysis, system design and appointment system.

An empirical study conducted by Creemer and Lamberecht. (2007) found that the capacity and variability analysis in a healthcare environment results in queuing models that are different from queuing model in industrial setting. He also showed the relationship between the capacity utilization, waiting time and patient (customer) service.

According to Biggs (2008) Elective surgery waiting lists are used to manage access to public hospital elective surgery services and give priority to those in most urgent need of care. They have become an integral feature of our health system, and allow limited health resources to be allocated or 'rationed' on the basis of need. Waiting lists also provide health consumers with an indication of how long they can expect to wait for their surgery.

According to Kolker (2008) the discrete event simulation model is more flexible; give more information than queuing analytic theory. Queuing theory is very volatile situation which cause unnecessary delay and reduce the service effectiveness of establishments. Apart from the time wasted, it is also leads breakdown of law and order.

Schoenmeyr et al. (2009) analysed that healthcare organizations function with very small net margins, so decisions about committing resources must be made with a high degree of confidence that the investment will lead to the desired result. The queuing approach is useful because it enables the investigation of future scenarios for which historical data are not directly applicable. Waiting times assist in measuring the rate of turnover on hospital waiting lists and are considered a more reliable indicator of hospital performance than the size of the waiting list. In some cases the patient may be removed from a waiting list. Reasons may include that they no longer require the procedure, are instead admitted as a emergency patient, receive their treatment at a different hospital or are transferred to the waiting list of a different hospital, are untraceable or die

Foster et al (2010) observed that Queuing models are useful in that they provide solutions to problems of waiting that are particularly relevant in health care. More generally, they illustrate the strengths of modeling in health care research and service delivery. Obamiro (2010) studied the waiting line for expectant women in Ante natal care unit. The results of the study evaluated the effectiveness of a queuing model in identifying the ante-natal queuing system efficiency parameters. He used TORA Optimization system to analyze data collected from ante-natal care unit of a public teaching hospital in Nigeria over a three-week period. The study showed that pregnant mothers spent less time in the queue and system in the first week than during the other succeeding two weeks.

Agrawal and Saxena (2010) analyzed the use of queuing theory in healthcare centre of IIT-K and the benefits accrued from the same and they conceptualized an appointment system in which customers who are about to enter service may have a probability of not being served and may re-join the queue. In their investigation, they found that the capacity of utilization is 76%, average number of people waiting in queue is calculated by the Poisson distribution method.

As examined by Mehandiratta (2011) with rapid change and alignment of health care system, new lines of services and facilities to render the same, server financial pressure on the health care organizations and extensive use of expanded managerial skills in healthcare setting, use of queuing models has become quite prevalent in it. Queuing models are used to achieve a balance or trade-off between capacity and service delays.

In the opinion of Enneyan (1997) healthcare continuous to become more competitive, the ability to assess trade-off between resources utilization, services, and operating costs grows in importance, such as with respect to appointment access, waiting delays and telephone services. More recently, simulation has been used to study and improve the quality of clinical laboratories, such as the diagnostic accuracy Pap smear, mammogram, HIV, Hepatitis result.

As studied by Szymanski (2003) the simulation provided quantifiable performance data which were used to input executive decision making. He focused on the simulation project approach which included five major phases:

- Develop conceptual model,
- Programming,
- Testing (Verification and Validation),
- Experimentation,
- Presentation.

Simulation to aid project leaders in advancing to the next level of sophistication with Six Sigma. Six Sigma Teams were created to review and analyze discrete sub-processes of the overall patient experience.

As it is analyzed by Hartvigsen (2004) SimQuick runs considerably more quickly as compared to others. This allows more simulations to be performed in a reasonable amount of time, which leads to more accurate results. As a result, more simulations are allowed. With process simulation, it can begin by building a computer model of a real world process. The initial goal is to have the computer model behave in a way similar to the real process, except much more quickly & various ideas for efficiency improvements on the computer model and use the best ideas on the real process.

Zhang et al (2000) has evaluated the performance of single-channel and multiple channels queues using the discrete-event simulation technique. The input to the simulators is based on live data. A customer can hop to a shorter queue but the service time needed by the customers in the queue may be longer thus resulting in an even longer waiting time.

In his study, Ahmed (2003) found that The Accident & Emergency Department is the dedicated area in a hospital that is organized and administered to provide a high standard of emergency care to those in community who perceived the need for or in need of acute or urgent care including hospitals admission.

Yeung et al. (2006) examined that the prioritization of treatment for patient with minor illness treatment or trauma over major patients with several illness can lead to the counter-intuitive outcome that mean response times for ambulance arrivals are not adversely affected.

Queuing theory uses models or mathematical models and performance measures to assess and hopefully improve the flow of customers through a queuing system (Bunday, 1996). A good patient flow means that the patient queuing is minimized while a poor patient flow means patients suffer considerable queuing delays. McClain (1976) reviews research on models for evaluating the impact of bed assignment policies on utilization, waiting time and the probability of turning away patients. Nosek and Wilson (2001) review the use of queuing theory in pharmacy applications with particular attention to improving customer satisfaction. Customer satisfaction is improved by predicting and reducing waiting times and adjusting staffing. Preater (2002) presents a brief history of the use of queuing theory in healthcare and points to an extensive bibliography of the research that lists many papers (however, it provides no description of the applications or results). Green (2006) presents the theory of queuing as applied in healthcare. She discusses the relationship amongst delays, utilization and the applications of the theory to determine the required number of servers.

Delays and queuing problems are most common features not only in our daily life situations such as at a bank or postal office, at a ticketing office, in public transportation or in a traffic jam but also in more technical environments, such as in manufacturing, computer networking and telecommunications. They play an essential role for business process re-engineering purposes in administrative tasks. 'Queuing models provide the analyst with a powerful tool for designing and evaluating the performance of queuing systems.' (Bank et. al. 2001). Whenever customers arrive at a service facility, some of them have to wait before they receive the desired service. It means that the customer has to wait for his/her turn, may be in a line. Customers arrive at a service facility (sales checkout zone in ICA) with several queues, each with one server (sales checkout counter). The customers choose a queue of a server according to some mechanism (e.g., shortest queue or shortest workload). (Adan, 2000) Sometimes, insufficiencies in services also occur due to an undue wait in service may be because of new employee. Delays in service jobs beyond their due time may result in losing future business opportunities. Queuing theory is the study of waiting in all these various situations. It uses queuing models to represent the various types of queuing systems that arise in practice. The models enable an appropriate balance between the cost of service and the amount of waiting.

Managing waiting lines create a great dilemma for managers seeking to improve the return on investment of their operations. Customers also dislike waiting for long time. If the waiting time and service time is high customers may leave the queue prematurely and this in turn results in customer dissatisfaction. This will reduce customer demand and eventually revenue and profit (Biju et. al. 2011). Though queuing analysis has been used in hospitals and other health-care settings, its use in this sector is not widespread. With rapid change and realignment of healthcare system, new lines of services and facilities to render the same, severe financial pressure on the healthcare organizations, and extensive use of expanded managerial skills in healthcare setting, use of queuing models has become quite prevalent in it. In an era of healthcare reform, improving quality and safety, and decreasing healthcare cost have become even more important goals than before. With rapid change and realignment of healthcare system, new lines of services and facilities to render the same, severe financial pressure on the healthcare organizations, and extensive use of expanded managerial skills in healthcare setting, The use of queuing model has become a prevalent analytical tool (Singh, 2007). Scientific management of patient flow is at the heart of our ability to achieve these goals. While on one hand we are faced with overcrowded facilities, on the other hand, the industry's financial conditions do not allow us to add resources liberally. One key challenge is our ability to match random patient demand to fixed capacity. Queuing theory is a methodology that addresses this very challenge. Queuing theory was first used in telecommunications and then was adopted by all major industries, like airlines, the Internet and most service-delivery organizations. In the health care industry, however, queuing theory has not been utilized until recently. When used appropriately, the results are often dramatic: saving time, increasing revenue, and increasing staff and patient satisfaction.

At its most basic level, queuing theory involves arrivals at a facility (i.e. computer store, pharmacy, bank) and service requirements of that facility (i.e., technicians, pharmacists, tellers). The number of arrivals generally fluctuates over the course of the hours that the facility is available for business. The theory enables mathematical analysis of several related processes, including arriving at the (back of the) queue, waiting in the queue (essentially a storage process), and being served at the front of the queue. The theory permits the derivation and calculation of several performance measures including the average waiting time in the queue or the system, the expected number waiting or receiving service, and the probability of encountering the system in certain states, such as empty, full, having an available server or having to wait a certain time to be served.

A good rule of thumb to remember the two distributions is that time between arrivals is exponentially distributed and the numbers of arrivals per unit of time is Poisson distributed.

The servicing or queuing system consists of the line(s) and the available number of servers. Factors to consider include the line length, number of lines and the queue discipline. Queue discipline is the priority rule, or rules, for determining the order of service to customers in a waiting line. One of the most commonly used priority rules is first-come-first-served, (FCFS). Others include a reservations first, treatment via triage (i.e., emergency rooms of health centres), highest-profit customer first, largest orders first, best customers first and longest wait-time first. An important feature of the waiting structure is the time the customer spends with the server once the service has started. This is referred to as the service rate: the capacity of the server in numbers of units per time period (i.e., 15 orders per hour).

There are two possible outcomes after a customer is served. The customer is either satisfied or not satisfied and requires re-service. The queue discipline refers to the order in which members of the queue are selected for service (Hillier and Lieberman, 1995). Winston and Albright (1997) states that the usual queue discipline is first come, first served (FCFS or FIFO), where customers are served in order of arrival. In this study the dental uses FCFS queuing discipline. Although, sometimes there are other service disciplines: last come, first served (which happens sometime in case of emergencies), or service-in-random order and priority rule. Davis et al, (2003) assert that reservations first, emergencies first, highest profit customer first, largest orders first, best customers first, longest waiting time in line, and soonest promised date are other examples of queue discipline. Unless otherwise stated, the queuing model adopted in this study assumes arrival from infinite source with infinite queue and with first in first served (FCFS) queue discipline, Obamiro, (2010)

According to Nosek and Wilson (2001), service mechanism describes howthe customer is served. In a single server system each customer is served by exactly one server,even though there may be multiple servers. In most cases, service times are random and they may vary greatly. Sometimes the service time may be similar for each job or constant. The service mechanism also describes the number of servers. A queuing system may operate with a single server or a number of parallel servers. An arrival who finds more than one free server may choose at random any one of them for receiving service. If he finds all the servers busy, he joins a queue common to all servers. The first customer from the common queue goes to the server who becomes free first (Medhi, 2003). A system may have an infinite capacity, that is, the queue in front of the server(s) may grow to any length. Furthermore, there may be limitation of space and so when the space is filled to capacity, an arrival will not be able to join the system and will be lost to the system. The system is called a delay system or a loss system, according to whether the capacity is infinite or finite respectively (Medhi, 2003). Once customers are served, they depart and may not likely re-enter the system to queue again. It is usually assumed that departing customers do not return into the system immediately (Adedayo, et al., 2006).

Economies around the world are more and more focussed on service industries in general and healthcare in particular (Parasuraman, 2004). In addition, patient flow times play an increasingly important role in today's healthcare systems. Government reimbursement systems (based on a justified length of stay), insurer's rejection of reimbursement (i.e. denied days), competition between hospitals, government regulations and patient satisfaction urge hospital decision makers to find ways to decrease waiting times (both waiting time inside the hospital as well as the waiting list that exists outside the hospital). Current healthcare literature and practice indicate that waiting lists and congested patient flows do indeed make up for one of the most important problems in care industries (Cerda and Rodriguez, 2006; Belson, 2006). In order to improve performance in an environment as complex as a hospital system, the dynamics at work need to be understood. To obtain such an understanding, queuing theory and simulation provide an ideal set of tools. Healthcare systems however have a number of specific features as compared to manufacturing systems, posing important methodological challenges. In healthcare, queuing theory has been used to assess capacity requirements (Kao Green, 2003; McManus et. al. 2004), outpatient scheduling Cayirli and Veral, (2003), process optimization (Vandaele et. al. 2003) and absence recovery modeling (Easton & Goodale, 2005).When the capacity of service provided fall short of the demand for the service.

Sanish (2007) in his article on application of queuing to the traffic at New Mangalore Port refers to queuing theory as an analytical techniques accepted as valuable tool for solving congestion problems. According to him the primary inputs to the models are the arrival and service patterns. These patterns are generally described by suitable random distribution. He observed that the arrival rate of ships follows exponential distribution while the service time follows Erlang or Poisson distribution. He observed that queuing theory can be used to predict some important parameters like average waiting time of ships, average queuing length, average number of ships in the port and average berth utilization factor closer to the actual values.

The effect of queuing during hospital visits in relation to the time spent for patients to access treatment in hospitals is increasingly becoming a major source of concern to a modern society that is currently exposed to great strides in technological advancement and speed (Stakutis and Boyle, 2009). Internal operational factors majorly determine outpatient waiting times and include; arrival pattern of prescriptions, sequencing of work, percentage of staff at work, interaction between the pharmacy service providers that is assessor and technician interaction or technician and counsellor interaction (Moss, 1987). A queue is a waiting line, whether of people, signals or things (Ashley, 2000). Queuing time is the amount of time a person, signal or thing spends before being attended to, or before value adding work is performed to or on it (Customer Management IQ, 2011). In many factories, queue time constitutes about 90% of the total lead time (Crabtree, 2008). Queue time can be extended to hospitals pharmacies where patients complain about the amount of time spent before treatment can be received (Zhang et al., 2000). Queues deal with problems which involve waiting for a product or service. Typical examples might be hospital/pharmacies/banks/supermarkets waiting for service, computers used in hospital management information system (HMIS) waiting for response, power failure situation that involves waiting for electric power to be restored to enable a piece of machinery used for compounding a prescription medication function. Another example is the public transport system- waiting for a train or a bus to convey a consultant pharmacist to the hospital in the advent of an accident, poisoning or hypertensive emergency. Queuing models have mainly been used to study congestion for interrupted traffic flows at signalized and unsignalized intersections (Woensel and Cruz, 2009). However, it has been shown that they can also be usefully applied to describe and analyse congestion for uninterrupted traffic flows in spite of the absence of the development of a formal queue. Queuing models applied to traffic situations provide an adequate description of the complex dynamic and stochastic environment under study (Woensel and Cruz, 2009). In designing queuing systems we need to aim for a balance in services rendered to clients. In essence, all queuing systems can be broken down into individual subsystems, consisting of entities queuing for some activity e.g. dispensing and counselling. Queue characteristics determines how, from the set of clients waiting for service, we choose the one to be served next (e.g. FIFO (first-in-first-out); LIFO (last-in-first-out); randomly). This order instituted for waiting is often called the queue discipline (David, 2005). Queue discipline can be examined so as to determine the queue characteristics to implore over a given period in providing pharmaceutical services from the outpatient section or from any other outlet (David, 2005; Noesk and Wilson, 2001). The queue discipline can include balking (clients or patients deciding not to join the queue if it is too long), reneging (clients or patients leave the queue if they have waited too long for service), jockeying (clients or patients switch between queues if they think they will get served faster by so doing).

Review of Waiting Line Model In Healthcare

As analysed by Fink & Gillett (2006) the cost of a dissatisfied customer is not negligible, they described Waiting in line is a primary source of dissatisfaction. They mentioned that a well-known queuing theories and integrating theory behind the Taguchi Loss Function, a manager can derive the costs associated with this dissatisfaction and that customer dissatisfaction is not just an issue at the upper specification limit, but rather for each moment in time beyond the targeted wait time. They illustrated by using the Taguchi Function, it can then be seen that these costs increase beyond the upper specification limit. However, by assessing these costs and then taking measures to reduce either the actual or perceived waiting times, organizations can quantitatively determine the cost-benefit relationship of improved waiting lines. Chin (2007), investigated the submittal review/approval process and used queuing theory to determine the major causes of long lead times. Under his study, he explored the underlying causes of waiting in a process flow and found the improvement methods from the queuing perspective. Yeung et al (2002), conducted research on large hospitals with the help of Laplace transform of the probability density function of customer response time in networks of queues with class-based priorities. He obtained the mean and standard deviation of total patient service time for large hospitals mainly for accident and emergency department. The widespread problem of extended waiting times for health services was examined by Georgiveskiy (2002) in the context of the Emergency Department (ED) at a regional hospital. A study of Georgievskiy et al (2002) examined Operations Research (OR) to reduce the waiting time in the hospital-admitting department. They conducted their study in five faces, from the collection of data to actual improvement in the quality of the health care delivery system. According to Koizumi (2002) the current trend toward downsizing and closing of state mental health institutions has led to an over-utilization of many local mental health facilities. This problem has often been intensifying by a shortage of long-stay psychiatric hospitals and community-type accommodations. He specifically examines blocking in interrelated mental health facilities, including :

(a) acute hospitals (where patients wait to enter extended acute hospitals),

(b) extended acute hospitals (where patients wait to enter residential facilities), and

(c) residential facilities (where patient wait to enter supported housing).

System has experienced severe congestion problems since 1992.

Singh (2006) found that the queuing theory in healthcare organizations is very beneficial. He used Queuing model to achieve a balance or trade-off between capacity and services delays & used the POM-QM Software for to demonstrate it. Fomundam et al. (2007) described the contributions and applications of queuing theory in the field of healthcare. They summarized a range of queuing theory results in areas of waiting time and utilization analysis, system design and appointment system. An empirical study conducted by Creemers et al. (2007) found that the capacity and variability analysis in a healthcare environment results in queuing models that are different from queuing model in industrial setting. He also showed the relationship between the capacity utilization, waiting time and patient (customer) service. According to Biggs (2008) Elective surgery waiting lists are used to manage access to public hospital elective surgery services and give priority to those in most urgent need of care. They have become an integral feature of our health system, and allow limited health resources to be allocated or 'rationed' on the basis of need. Waiting lists also provide health consumers with an indication of how long they can expect to wait for their surgery.

According to Kolker (2008) the discrete event simulation model is more flexible; give more information than queuing analytic theory. Queuing theory is very volatile situation which cause unnecessary delay and reduce the service effectiveness of establishments. Apart from the time wasted, it is also leads breakdown of law and order. Many lives and property had been lost in queues at filling stations in past. (Adeleke, Ogunwala, Halid 2009). Schoenmeyr et al. (2009) analysed that healthcare organizations function with very small net margins, so decisions about committing resources must be made with a high degree of confidence that the investment will lead to the desired result. The queuing approach is useful because it enables the investigation of future scenarios for which historical data are not directly applicable. Waiting times assist in measuring the rate of turnover on hospital waiting lists and are considered a more reliable indicator of hospital performance than the size of the waiting list. In some cases the patient may be removed from a waiting list. Reasons may include that they no longer require the procedure, are instead admitted as a emergency patient, receive their treatment at a different hospital or are transferred to the waiting list of a different hospital, are untraceable or die. Foster et. al. (2010) observed that Queuing models are useful in that they provide solutions to problems of waiting that are particularly relevant in health care. More generally, they illustrate the strengths of modelling in health care research and service delivery. Obamiro (2010) studied the waiting line for expectant women in Ante natal care unit. The results of the study evaluated the effectiveness of a queuing model in identifying the ante-natal queuing system efficiency parameters. He used Tora Optimization system to analyze data collected from ante-natal care unit of a public teaching hospital in Nigeria over a three-week period. The study showed that pregnant mothers spent less time in the queue and system in the first week than during the other succeeding two weeks. Agrawal and Saxena (2010) analyzed the use of queuing theory in healthcare centre of IIT-K and the benefits accrued from the same and they conceptualize an appointment system in which customers who are about to enter service may have a probability of not being served and may re-join the queue. In their investigation, they found that the capacity utilization is 76%, average number of people waiting in queue is calculated by the Poisson distribution method. As examined by Mehandiratta (2011) with rapid change and alignment of health care system, new lines of services and facilities to render the same, server financial pressure on the health care organizations and extensive use of expanded managerial skills in healthcare setting, use of queuing models has become quite prevalent in it. Queuing models are used to achieve a balance or trade off between capacity and service delays. In the opinion of Enneyan (1997) healthcare continuous to become more competitive, the ability to assess trade-off between resources utilization, services, and operating costs grows in importance, such as with respect to appointment access, waiting delays and telephone services. More recently simulation has been used to study and improve the quality of clinical laboratories, such as the diagnostic accuracy Pap smear, mammogram, HIV, Hepatitis result. As studied by Szymanski (2003) the simulation provided quantifiable performance data which were used to input executive decision making. He focused on the simulation project approach which included five major phases:

- Develop conceptual model,
- Programming,
- Testing (Verification and Validation),
- Experimentation,
- Presentation.

Simulation to aid project leaders in advancing to the next level of sophistication with Six Sigma. Six Sigma Teams were created to review and analyze discrete sub-processes of the overall patient experience. As it is analyzed by Hartvigsen (2004) SimQuick runs considerably more quickly as compared to others. This allows more simulations to be performed in a reasonable amount of time, which leads to more accurate results. As a result, more simulations are allowed. With process simulation, it can begin by building a computer model of a real world process. The initial goal is to have the computer model behave in a way similar to the real process, except much more quickly & various ideas for efficiency improvements on the computer model and use the best ideas on the real process.

Zhang et al. (2000) has evaluated the performance of single-channel and multiple channels queues using the discrete-event simulation technique. The input to the simulators is based on live data. A customer can hop to a shorter queue but the service time needed by the customers in the queue may be longer thus resulting in an even longer waiting time. In his study, Ahmed (2003) found that The Accident & Emergency Department is the dedicated area in a hospital that is organized and administered to provide a high standard of emergency care to those in community who perceived the need for or in need of acute or urgent care including hospitals admission. Yeung et al. (2006) examined that the prioritization of treatment for patient with minor illness treatment or trauma over major patients with several illness can lead to the counter-intuitive outcome that mean response times for ambulance arrivals are not adversely affected. As examined by Kumar & Shim (2008) the (BPR) Business Process Reengineering is essential to improve the efficiency in Emergency Department.

Chapter 3

METHODOLOGY

3.1 Introduction

This chapter looks at and gives report on the population from which the sample was drawn, instruments and procedures used for the collection and analysis of the data. The overall model structure is explained and the methodology used to achieve the objectives under study is clearly state.

3.2 Sources and The Type of Data

The data for this study were collected from the dental department of the Essikado hospital. The data is purely from a primary source. Patients coming for dental treatment were not made to join the queue at the out-patient department. They were quickly attended to and made to join the queue at the dental section. This is because the dentist closes by 1.00 pm. The data of most fundamental importance were arrival time, service time and departure time of patients. These data were collected within a three hour interval.

3.3 Data Collection Procedure

The researcher was taken through all the processes at the dental section. Data was collected over a period of one week from 8:30 am to 11:30 am each day. The research goal during the collection of data was to get time between patient consecutive arrivals and service length. More interest was taken in periods where each of the following events occurs: a patient arrival, a beginning of service and an end of service in the dental clinic.

3.4 Queuing Model and Kendall's Notation

The basic queuing model is shown below. It can be used to model for example, machine or operators processing orders or communication equipment processing information. Among others, a queuing model is characterized by:

a. The arrival process of customers: Usually we assume that the inter-arrival times are independent and have a common distribution. In many practical situations customers arrive according to a Poisson stream (that is, exponential inter-arrival times). Customers may arrive one by one, or in batches. An example of batch arrivals is the customs office at the border where travel documents of bus passengers have to be checked.

b. The behaviour of customers: Customers may be patient and willing to wait (for a long time), or customers may be impatient and leave after a while. For example, in call centres, customers will hang up when they have to wait too long before an operator is available, and they possibly try again after a while.

c. The service times: Usually we assume that the service times are independent and identically distributed, and that they are independent of the inter-arrival times. For example, the service times can be deterministic or exponentially distributed. It is also possible that service times are dependent of the queue length. For example, the processing rates of the machines in a production system can be increased once the number of jobs waiting to be processed become too large.

d. The service discipline: Customers can be served one by one or in batches. We have many possibilities for the order in which they enter service:

   i. first come first served, i.e. in order of arrival

   ii. random order;

iii. last come first served (e.g. in a computer stack or a shunt buffer in a production line);

iv. priorities (e.g. rush orders first, shortest processing time first);

v. processor sharing (in computers that equally divide their processing power over all jobs in the system).

e. The service capacity: There may be a single server or a group of servers helping the customers.

f. The waiting room: There can be limitations with respect to the number of customers in the system. For example, in a data communication network, only finitely many cells can be buffered in a switch. The determination of good buffer sizes is an important issue in the design of these networks.

Kendall introduced a shorthand notation to characterize a range of these queueing models. It is a three-part code a/b/c. The first letter specifies the interarrival time, the second one the service time distribution and the third signifies the number of servers. For example, a/b/1 is a queuing model with a single server.

In the basic model, customers arrive one by one and they are always allowed to enter the system, there is always room, there are no priority rules and customers are served in order of arrival. It will be explicitly indicated (e.g. by additional letters) when one of these assumptions does not hold.

3.4.1 Occupation Rate

In a single-server system G/G/1 with arrival rate, λ and mean service time $E(B)$ the amount of work arriving per unit time equals $\lambda E(B)$. The server can handle 1 unit work per unit time. To avoid that the queue eventually grows to infinity, it require that $\lambda E(B) < 1$. It is common to use the notation

$$\rho = \lambda E(B)$$

If $\rho < 1$, then ρ is called the occupation rate or server utilisation, because it is the fraction of time the server is working.

In a multi-server system G/G/c it requires that $\lambda E(B) < c$. Here the occupation rate per server is $\rho = \lambda E(B)/c$.

3.4.2 Performance Measure

Relevant performance measures in the analysis of queueing models are:

- The distribution of the waiting time and the sojourn time of a customer. The sojourn time is the waiting time plus the service time.
- The distribution of the number of customers in the system (including or excluding the one or those in service)
- The distribution of the amount of work in the system. That is the sum of service times of the waiting customers and the residual service time of the customer in service.
- The distribution of the busy period of the server. This is a period of time during which the server is working continuously.

The aim of all investigations in queuing theory is to get the main performance measures of the system which are the probabilistic properties (distribution function, density function, mean, variance) of the following random variables: number of customers in the system, number of waiting customers, utilization of the server/s, response time of a customer, waiting time of a customer, idle time of the server, busy time of a server. Of course, the answers heavily depends on the assumptions concerning the distribution of inter-arrival times, service times, number of servers, capacity and service discipline. It is quite rare, except for elementary or Markovian systems, that the distributions can be computed. Usually their mean or transforms can be calculated.

3.5 Formulation of Methods

3.5.1 The mean arrival rate

Let λ be the mean arrival rate and let n be the number of patients that entered the system between 8:30 am - 11:30 am. Also, let h be the number of hours between 8:30 am - 11:30 am. Then, the mean arrival rate is given by the formula,

$$\lambda = \frac{n}{h} \qquad \text{arrival per hour} \tag{3.1}$$

3.5.2 The mean service rate

The mean service rate at the dental section

Let $S_1, S_2, \ldots, S_n$ be the observed service times when patients arrived at the hospital. Let n be the number of patients that are attended to by the dentist and let b be the start service time for each patient and let e be the finished service time for each patient. From these defined letters

$$\begin{aligned} S_1 &= \frac{e_1 - b_1}{n_1} \\ S_2 &= \frac{e_2 - b_2}{n_2} \\ &\vdots \\ S_n &= \frac{e_n - b_n}{n_n} \end{aligned}$$

Little's Formula

The Little's formula is given by:

$$L = \lambda W \tag{3.2}$$

This formula is used to determine the amount of time that a patient would spend in the system. From this formula, L is defined as the average number of patients present in the queuing system and λ is defined as the mean arrival rate. W is also defined as the expected time a patient spends in the queuing system. We could say from Little's formula that:

If equation (3.2) is Little's second flow of equation. then

$$W_q = \frac{L_q}{W} \tag{3.3}$$

Equation (3.5) which follows directly from Little's second flow equation was used for the single-channel and the multiple-channel waiting line models.

The general expression that applies to waiting line models is that the average time in the system W, is equal to the average time, W_q plus the average service time. A system with a mean service rate μ, the average or mean service time is $\frac{1}{\mu}$. Hence, we have the following general relationships:

$$L_q = \lambda W_q \qquad \text{and} \tag{3.4}$$

$$L_s = \lambda W_s \tag{3.5}$$

Where equations (3.5) and (3.6) are defined as the average number of patients in the waiting line and the average number of patients in the system respectively where

W_q = Expected waiting time in queue

W_s = Expected time a customer spends in the system

L_q = Expected number of customers in the queue

L_s = Expected number of customers in the system.

3.6 The Operating Characteristics of M/M/I Model (Single Server - Infinite Population)

$P = \frac{1}{\mu}$ is defined as the average utilization of the system i.e the probability that the system is

$$P(\text{n customers during period T}) = \frac{e^{-\lambda t}(\lambda t)^n}{n!} \tag{3.6}$$

When the time taken to serve different customers is independent then;

$$P(\text{not more than T time period needed to serve a customer}) = 1 - e^{\mu T} \tag{3.7}$$

1. Probability that the system is idle i.e the probability that there are no customers in the system;

$$P_o = 1 - P = 1 - \frac{\lambda}{\mu} \tag{3.8}$$

2. The probability of having exactly n customers in the system

$$P_n = P^n P_o = 1 - \frac{\lambda}{\mu} = P^n(1 - P) \tag{3.9}$$

3. Expected number of customers in the system

$$L_s = \frac{\lambda}{\mu - \lambda} \tag{3.10}$$

OR

$$\frac{P}{1 - P} \tag{3.11}$$

4. Expected number of customers in the queue

$$L_q = L_s - \frac{\lambda}{\mu} = \frac{\lambda}{\mu - \lambda} - \frac{\lambda}{\mu} = \frac{\lambda^2}{\mu(\lambda - \mu)} = \frac{P^2}{1 - P} \tag{3.12}$$

Equation (3.13) can also be noted as the average length of all queues including empty queues.

5. The average length of non empty queues (i.e. those which contain at least one customer)

$$L_n = \frac{\mu}{\mu - \lambda} = \frac{1}{1 - P} \tag{3.13}$$

6. Expected waiting time in the queue

$$W_q = \frac{1}{\lambda} - L_q \tag{3.14}$$

Substituting equation (3.11) into (3.15), We have

$$W_q = \frac{\lambda}{\mu(\mu - \lambda)} = \frac{P}{\mu - \lambda} \tag{3.15}$$

7. Expected time a customer spends in the system

$$W_s = \frac{1}{\lambda} \times L_s \tag{3.16}$$

Putting equation (3.11) into equation (3.17), we have;

$$W_q = \frac{1}{\mu - \lambda} \tag{3.17}$$

Since the mean service rate is ?, the average (expected) time for completing the service is $\frac{1}{\mu}$

Therefore, the expected time a customer would spend in the system would be equal to the expected waiting time in the queue plus the average servicing time. Thus,

$$W_s = W_q + \frac{1}{\mu} = \frac{\lambda}{\mu(\mu - \lambda)} + \frac{1}{\mu} = \frac{1}{\mu - \lambda} \tag{3.18}$$

Which is the same as equation (3.18), as shown earlier

8. The probability that a customer spends more than t units of time in the

system

$$W_s(t) = e^{\frac{-t}{s}} \tag{3.19}$$

9. The probability that a customer spends more than t units of time in the queue

$$W_q = Pe^{-\frac{t}{W_s}} \tag{3.20}$$

EXAMPLE

A local shop has a single counter at which customers are served. Customers arrive at a rate of 20 per hour according to a Poisson distribution and service times are exponentially distributed with a mean rate of 24 customers per hour.

(a) What is the mean number of customers waiting in the queue?

(b) What is the mean time a customer spends in the queue?

(c) What is the mean number of customers in the queue?

(d) What is the average time a customer spends from arrival until fully served.

(e) Calculate the server utilization.

Solution

$\lambda = 20 \qquad \mu = 24$

(a) L_q(mean number of customers in the queue) $= \lambda W_q = 20(0.2083) = 4.166$ i.e 4 customers

(b)

$$W_q(\text{mean time it takes a customer to start being served}) = \frac{\rho}{\mu - \lambda}$$

$$W_q = \frac{0.8333}{24 - 20} = \frac{0.8333}{4} = 0.2083\text{hours}$$

(c)

$$L_s(\text{mean number of customers in the system}) = \lambda W_s = 20(0.24996) = 5customers$$

(d)

$$W_s(\text{mean time spent by a customer from arrival until fully served}) = W_q + \frac{1}{\mu}$$

$$W_s = 0.2083 + \frac{1}{24} = 0.2500\,\text{hours}$$

(e)

$$\rho(\text{server utilization: percentage of time a server is being utilized by a customer}) = \frac{\lambda}{\mu}$$

$$P = \frac{20}{24} = 0.8333$$

3.7 Single Server Model - Finite Population

This model is based on similar assumptions as that of M/M/1 model except that the input population is finite. For this model, the system structure is such that we have a total of M customers, a customer is either in the system (consisting of a queue and a single server) or outside the system and in a sense, arriving. When a customer is in the arriving condition, then the time it takes him to arrive is a random variable having an exponential distribution with mean equal to $\frac{1}{\lambda}$. When there are n customers in the system, then there is M ? n customers in the arriving state. From this, the total average rate of arrivals in the system is $\lambda(M - n)$.

The single server model with finite population is self-regulating. This means that when the system gets busy, with many customers in the queue, then the rate at which additional customers arrive is reduced thus lowering the con-

gestion in the system. In this model there is a depending relationship between arrivals. For the reason of dependency relationships between arrivals, the Poisson probability distribution law cannot be strictly applied when the input population is finite. Instead of the arrivals' statement as an average for the population, we classify them as an average of a unit time.
Hence, the exponential distribution with mean $= \frac{1}{\lambda}$

3.8 The Operating Characteristics of a Single Server - Finite Population Model

The number of customers in the source population = M.
Average inter-arrival between successive arrivals $= \frac{1}{\lambda}$
Service rate $= \lambda$

1. Probability that the system would be idle;

$$P_o = \left[\sum_{i=0}^{m} \left(\frac{M!}{(M-1)!} \left(\frac{\lambda}{\mu} \right)^t \right) \right]^{-1} \tag{3.21}$$

2. Probability of n customers in the system

$$P_n = \left[P_o \left(\frac{\lambda}{\mu} \right)^n \frac{M!}{(M-n)!} \right], 0 < n \leq M, or\ n > M \tag{3.22}$$

3. Expected length of the queue.

$$L_q = M - \frac{\lambda + \mu}{\mu} (1 - P_o) \tag{3.23}$$

4. Expected number of customers in the system,

$$L_s = L_q + (1 - P_o) = M - \frac{\mu}{\lambda} (1 - P_o) \tag{3.24}$$

5. Expected waiting time of a customer in the queue,

$$W_q = \frac{L_q}{\mu(1 - P_o)} = \frac{1}{\mu}\left(\frac{M}{(1 - P_o)} - \frac{(\lambda + \mu)}{\lambda}\right) \tag{3.25}$$

6. Expected time a customer spends in the system,

$$W_s = W_q + \frac{1}{\lambda} = \frac{1}{\mu}\left[\frac{M}{(1 - P_o)} - \frac{(\lambda + \mu)}{\lambda} + 1\right] \tag{3.26}$$

3.9 The Operating Characteristics of a Multiple Server Model (M/M/K/∞)

The model is based on the following assumptions;

(i) The arrival of customers follows Poisson probability distribution

(ii) The service time has an exponential distribution

(iii) There are k servers, each of which provides identical services

(iv) A single waiting line is formed

(v) The input population is infinite

(vi) The service is on a first-come-first-served basis

(vii) The arrival rate is smaller than the combined service rate of all k service facilities

3.10 Operating Characteristics of a Single Server - Finite Population Model

$$\text{Average rate of arrivals} = \lambda \tag{3.27}$$

$$\text{Numbers of servers} = k \tag{3.28}$$

$$\text{Mean combined rate of all servers} = k\mu \tag{3.29}$$

$$\text{Utilization factor of the entire system} - P = \frac{\lambda}{k\mu} \tag{3.30}$$

Probability that the system shall be idle

1. $$P_o = \left[\sum_{i=0}^{k-1} \frac{(\frac{\lambda}{\mu})^t}{i!} + \frac{(\frac{\lambda}{\mu})^k}{k!(1-P)}\right]^{-1} \tag{3.31}$$

   The probability that there would be exactly n customers in the system,

2. $$P_n = (P_o)\frac{\left(\frac{\lambda}{\mu}\right)^n}{n!}, \text{when } n \leq k \tag{3.32}$$

   And

   $$P_n = (P_o)\left(\frac{\left(\frac{\lambda}{\mu}\right)^n}{k!k^{n-1}}\right), \text{ when } n > k \tag{3.33}$$

3. The expected number of customers in the waiting line

   $$L_q = \frac{\left(\frac{\lambda}{\mu}\right)^k P}{k!(1-P)^2}(P_o) \tag{3.34}$$

4. The expected number of customers in the system,

   $$L_s = L_q + \frac{\lambda}{\mu} \tag{3.35}$$

5. The expected waiting time in the queue

   $$W_q = \frac{L_q}{\lambda} \tag{3.36}$$

6. The expected time a customer spends in the system,

   $$W_s = W_q + \frac{1}{\lambda} \tag{3.37}$$

EXAMPLE

Customers queue in a single line in a department store and are served at one of 3 tills on a first-come first-served basis. Customers arrive at a rate of 24 per hour (according to a Poisson distribution) and service times are exponentially distributed with a mean rate of 10 customers per hour.

(a) What is the mean number of customers waiting in the queue?

(b) What is the mean time a customer spends in the queue?

(C) What is the mean number of customers in the queue?

(d) What is the average time a customer spends from arrival until fully served.

(e) Calculate the server utilization

Solution

$\lambda = 24$ $\mu = 10$ k(numbers of servers)=3

$$P_o = 1/((24/10)^o/0! + (24/10)^1/1! + (24/10)^2/2!)$$
$$= 1/(1 + 2.4 + 2.88) = 0.16$$

(a)

$$L_q = (24 * 10 * (24/10)3/((3-1)! * (3*10-24)2)) * 0.16$$
$$= (3317.76/72) * 0.16$$
$$= 7.37 customers$$

(b)

$$W_q(\text{mean time it takes a customer to start being served}) = L_q/\lambda$$

$$W_q = \frac{7.37}{24} = 0.31\text{hours}$$

(c)

$$L_s(\text{mean number of customers in the system}) = L_q + \frac{\lambda}{\mu}$$
$$= 7.37 + 24/10 = 9.77$$

(d)

$$W_s(\text{mean time spent by a customer from arrival until fully served}) = W_q + \frac{1}{\lambda}$$

$$W_s = 0.31 + 1/24 = 0.01292\text{hours}$$

(e)

e) Î¡ (server utilization: percentage of time a server is being utilized by a customer) $= \frac{\lambda}{k\mu}$

$$P = \frac{24}{3(10)} = 0.8$$

3.11 The Erlang Distribution

The Erlang distribution is used to model situations where the inter-arrival times do not appear to be exponential. It is a continuous random variable (T) whose density function is specialized by two parameters; a scale parameter, ?, and a shape parameter k (where k is a positive integer).

A shape parameter and a scale parameter are kinds of a numerical parameter of a parametric family of probability distributions. The larger the scale factor, the more spread out the distribution. The scale parameter is useful in modelling applications since they are flexible enough to model a variety of data sets. A shape parameter allows a distribution to take on a variety of shapes, depending on the value of the shape parameter.

The Erlang distribution $(E_K(\mu))$ is given by;

$$F(t) = \frac{\mu(\mu t)^{k-1}}{(k-1)!}e^{-\mu t},\ t > 0 \tag{3.38}$$

The distribution function equals

$$F(t) = 1 - \sum_{j=0}^{k-1} \frac{(\mu t)^{k-1}}{j!}e^{-\mu t},\ t > 0 \tag{3.39}$$

Applying integration by parts, the mean, variance and squared coefficients of variation are equal to;

$$E(T) = \frac{k}{\mu}, \qquad \sigma^2 = \frac{k}{\mu}, \qquad C^2(T) = \frac{1}{K}$$

3.12 Poisson Process

The Poisson process is an extremely useful process for modelling purposes in many practical applications such as, e.g.; to model arrival process for queuing models or demand processes for inventory systems. It could be used to approximate stochastic processes. In the Poisson probability distribution, the observer records the number of events in a time interval. The Poisson process is a continuous-time process.

Let $N(t)$ be the number of arrivals in $[0,t]$ for a Poisson process with rate, λ, i.e. the time between successive arrivals is exponentially distributed with parameter, λ, and independent of the past. Then, the Poisson distribution with parameter, λt for $N(t)$ can be calculated as;

$$P[N(t) = k] = \frac{(\lambda t)^k}{k!}e^{-\lambda t},\ k = 0, 1, 2, \ldots \tag{3.40}$$

Applying integration by parts gives

$$E[N(t)] = \lambda t, \qquad \sigma^2[N(t)] = \lambda t$$

The mean of the Poisson Process is the same as the variance. (Wayne,1991).

EXAMPLE

A life insurance salesman sells on the average 3 life insurance policies per week. Use Poisson's law to calculate the probability that in a given week he will sell some policies.

Here, $\mu = 3$

'Some policies' means '1 or more policies'. We can work this out by finding 1 minus the 'zero policies' probability:

$$P(X > 0) = 1 - P(X_o)$$

$$\text{Now } P(X) = \frac{e^{-\mu}\mu^x}{x!} \quad \text{so } P(X_o) = \frac{e^{-3}3^0}{0!} = 4.9787 \times 10^{-2}$$

Therefore the probability of 1 or more policies is given by:

$$\text{Probability} = P(X \geq 0) = 1 - P(X_o) = 1 - 4.9787 \times 10^{-2} = 0.95021$$

3.13 Pure Birth Model and Death Models

The pure birth-death process is used to study the number of customers in a queue. It is a special case of continuous stochastic process where the state transitions are of only two types: ?births? which increase the state variable by one and ?deaths? which decrease the state by one.

When a birth occurs, the process goes from state n to $n + 1$. When a death occurs, the process goes from n to state $n - 1$. The process is specialized by birth rates

$$(\lambda_i), \quad i = 0, 1, 2, 3, \ldots \infty$$

and death rates

$$(\mu_i), \quad i = 1, 2, 3, \ldots \infty$$

Poisson process is a pure birth process where $\lambda_i = \lambda$ for all $i \geq 0$

Birth process is the same as arrival process and death process is the departure process.

The birth-death process describes probabilistically how the number of customers, $N(t)$ in the queuing system changes as time (t) increases.

The birth-death process is governed by these assumptions:

(i) Given $N(t) = n$, the current probability distribution of the remaining time until the next birth (arrival) is exponential with parameter λ_n.

(ii) Given $N(t) = n$, the current probability distribution of the remaining time until the next death (service completion) is exponential with parameter μ_n.

(iii) The random variable of assumptions one and the random variable of assumption two are mutually independent. The next transition in the state of the process is either $n \longrightarrow (n+1)$ or $n \longrightarrow (n-1)$ depending on whether the former or the latter random variable is smaller.(Wayne, 1991).

3.14 The Traffic Intensity (ρ)

An important parameter in any queuing system is the traffic intensity also called the load or the utilization, defined as the ratio of the mean service time $E(X) = \frac{1}{\mu}$ over the mean inter - arrival time $E(\tau) = \frac{1}{\lambda}$

$$\rho = \frac{E(x)}{E(\tau)} = \frac{\lambda}{\mu} \tag{3.41}$$

Where λ and μ are the mean inter-arrival and service rate, respectively.

Clearly, if $\rho > 1$ or $E[x] > E[\tau]$, which means that the mean service time is longer than the mean inter-arrival time, then the queue will grow indefinitely long for large t, because packets are arriving faster on average than they could be served. In this case $(\rho > 1)$, the queuing system is unstable or will never reach a steady-state. The case where $\rho = 1$ is critical. In practice, therefore, mostly situations where $\rho < 1$ are of interest.

If $\rho < 1$ a steady state can be reached. The considerations are a direct consequence of the law of conservation of packets in the system.

3.15 Queuing Model

3.15.1 Kendall's notation

In the year 1953, Kendall introduced a notation that is commonly used to describe and classify the type of a queuing model that a queuing system corresponds to. The general syntax is A/B/n/K/m, where A specifies the inter-arrival process, B the service process, n the number of servers, K the number of positions in the queue and m restricts the number of allowed arrivals in the queuing system. Examples for both the inter-arrival distribution A and the service distribution B are M (memoryless or Markovian) for the exponential distribution, G for a general distribution and D for a deterministic distribution.

When other letters are used besides these three common assignments, the meaning will be defined. For example, M/G/1 stands for a queuing system with exponentially distributed Inter-arrival times, a general service distribution and 1 server. If one of the two last identifiers K and m is not written, they should be interpreted as infinity. Hence, M/G/1 has an infinitely long queue and no restriction on the number of allowed arrivals.

A queue is therefore, described in a shorthand notation by A/B/C/K/N/D or the more concise A/B/C. in the concise version, it is assumed that $K = \infty$, $N = \infty$ and D = FCFS (first come, first served).

For Kendall's notation, the letters A and B can also be described by

Table 3.1: Description of Kendall's notation

Letter	Description
A	The arrival process
B	The service time distribution
C	The number of servers
K	The number of channels in the system
N	The calling population
D	The queue discipline

Table 3.2

Table 3.2: Examples of queuing models

Letter	Description
M	Exponential distribution
D	Deterministic inter-arrival times
E^K	Erlang distribution (k=shape parameter)
G	General distribution

3.15.2 Examples of queuing models

(i) **M/M/1 Model**

In this model, random arrivals and exponentially distributed service times are assumed. Poisson distribution is used to define the random arrivals. The M/M/1 Model has only one server serving customers on first come, first served bases. The population here is infinite, so arriving customers are unaffected by the queue size. The parameters given for M/M/1 model are λ (the average arrival rate), μ (the average service rate) which may be calculated from the average service time.

(ii) **M/M/M/∞**

This describes a queuing system with M number of servers and infinite number of waiting lines.

(iii) **M/M/1/∞/K**

This describes a queuing system with a single server, infinite number of waiting lines and finite customer population k.

(iv) **M/E^2/2/K**

This describes a queuing system with Poisson arrivals, Erlangian of order 2 service time distribution, 2 servers, and maximum number of k in a queue.

(v) **M/D/5/40/200/FCFS**

Where M is the exponential distributed times, D is the deterministic service times, 5 is the number of servers, 40 is the number of buffers (35 for waiting), a total population of 200 customers and FCFS is the service discipline (first come, first served).

3.16 Deterministic Queuing Models

Queuing models can be categorized as being deterministic if customers arrive at known intervals and the service time is known with certainty. Suppose that customers come to bank's teller counter every 5 minutes. Thus the interval between the arrivals of any two successive customers in exactly 5 minutes, suppose that the banker takes exactly 5 minutes to serve each customer. Here the arrival and the service rates are each equal to 12 customers per hour. That is, 60 minutes/5 minutes = 12 customers per every 5 minutes. In this situation, there shall be never a queue and the banker shall always be busy with work.

Now, suppose that the banker can serve 15 customers per hour, the consequence of this higher service rate would be that the banker would be busy 4/5th of the time and idle in 1/5th of the time. The teller shall take 4 minutes to serve a customer and wait for the next customer to come. Here also, there shall never be a queue as before.

If on the other hand, the banker can serve only 10 customers per hour, then the result would be that he would be always busy and the queue length will increase continuously without limit with the passage of time. It is easy to see when the service rate is less than the arrival rate, the service facility cannot cope with all the arrivals and eventually the system leads to explosive situations. This problem can be resolved by providing additional service station(s).

Symbolically let λ be the arrival rate of customers per unit time and μ be the service rate per unit time. Then,

If $\lambda > \mu$, the waiting line shall be formed which will increase indefinitely; the service facility would always be busy; and the service system will eventually fail.

If $\lambda \geq \mu$, there shall be no waiting time; the proportion of time the service facility would be idle is $1 - \lambda/\mu$. Where $\lambda/\mu = P$ is called the average utilization or the traffic ratio, or the clearing ratio. This indicates the proportion of time or the probability that the service station is busy.

From this model,

If $P > 1$, the arrivals come at a faster rate than the server can accommodate. The expected queue increases without limit and a steady state does not occur. This would make the system to fail ultimately.

If $P \geq 1$, the system would work and P is the utilization factor of the system, that is, the proportion of time the system busy. Also, if $P < 1$, then steady-state probabilities would occur.

The deterministic queuing model may exist when we are dealing with, for example movements of items for processing in a highly automated plant. However, generally and more particularly when human beings are involved, the arrivals and servicing time are variable and uncertain. Thus, variable arrival rates and servicing times are the more realistic assumptions. The probabilistic queuing models are based on these assumptions.

3.17 Summary

In this chapter, the research methodology adopted for the study, sources and type of data collection procedure were discussed. Formulation of methods, the operating characteristics of queuing models and probability distributions behind queuing theory were put across.

Analysis of the collected data is presented in the next chapter.

Chapter 4

COLLECTION AND ANALYSIS OF DATA

4.1 Introduction

This chapter looks at the analysis and the modelling of the data collected. Tables were used to show the data collected and the findings were used to answer the research questions that were formulated in chapter one.

4.2 Source and Type of Data

The data for this study were collected from the dental department of the Essikado hospital. The data was purely from a primary source. Patients coming for dental treatment are not made to join the queue at the out-patient department. They were quickly attended to and made to join the queue at the dental section. The data of most fundamental importance were arrival time, service time and departure time of patients. These data were collected within a three hour interval.

4.3 Data Collection Procedure

The researcher was taken through all the processes at the dental section. Data were collected over a period of two weeks from 8:30 am to 11:30 am each day. The research goal during the collection of data was to get time between patient consecutive arrivals and service length. More interest was taken in periods where each of the following events occurred: patient arrival, beginning of service and end of service in the dental department.

4.4 Data Collection

The following were the data collected for each day.

4.4.1 Presentation of Data Collection from 13th to 17th 2015

Table 4.1: Presentation of data collected on Monday

No. of customers	Arrival Time	Service Start Time	Service End Time	Service Time
1	08:38	10:53	10:59	00:06
2	08:49	10:55	11:14	00:19
3	09:02	10:58	11:16	00:18
4	09:31	11:01	11:25	00:24
5	09:32	11:20	11:32	00:12
6	09:47	11:30	11:35	00:05

Table 4.1 shows the number of patients that arrived at the dental department on Monday during the data collection. The first column shows the patient number while the second and third column shows the arrival time and the service start time respectively. The fourth and fifth column shows the service end time and the service times respectively. Six patients were recorded on Monday.

Table 4.2: Presentation of data collected on Tuesday

No. of customers	Arrival Time	Service Start Time	Service End Time	Service Time
1	08:32	10:10	10:30	00:20
2	08:36	10:16	10:35	00:19
3	08:40	10:36	10:40	00:04
4	09:00	10:41	10:59	00:18
5	09:10	11:05	11:25	00:20

From Table 4.2 five patients were recorded on Tuesday. The columns show the patient number, the arrival time, service start time, service end time and the service time. The first and fifth patients recorded the highest service time.

Table 4.3: Presentation of data collected on Wednesday

No. of customers	Arrival Time	Service Start Time	Service End Time	Service Time
1	08:40	10:46	11:00	00:14
2	08:52	10:49	11:05	00:16
3	09:39	10:50	11:13	00:23
4	10:14	10:54	10:59	00:05
5	10:17	11:16	11:21	00:05

Table 4.3 shows the number of patients that arrived at the dental department on Monday during the data collection. The first column shows the patient number while the second and third column shows the arrival time and the service start time respectively. The fourth and fifth column shows the service end time and the service times respectively. Five patients were recorded on Wednesday.

Table 4.4: Presentation of data collected on Thursday

No. of customers	Arrival Time	Service Start Time	Service End Time	Service Time
1	08:39	09:56	10:11	00:15
2	08:48	09:59	10:13	00:14
3	08:55	10:02	10:08	00:06
4	10:15	10:22	10:26	00:04

Table 4.4 shows the number of patients that arrived at the dental department on Thursday during the data collection. The first column shows the patient number while the second and third column shows the arrival time and the service start time respectively. The fourth and fifth column shows the service end time and the service times respectively. Thursday recorded four patients

From Table 4.5, four patients were recorded on Friday. The columns show the patient number, the arrival time, service start time, service end time and the service time. The first and fifth patients recorded the highest service time.

Table 4.5: Presentation of data collected on Friday

No. of customers	Arrival Time	Service Start Time	Service End Time	Service Time
1	08:33	09:44	10:11	00:20
2	08:37	09:47	10:13	00:18
3	08:45	09:50	10:09	00:19
4	09:17	09:57	10:12	00:15

4.5 Analysis of Data Collected

This section deals with the analysis of data collected from the dental department at Essikado Hospital, Essikado. The mean arrival rates and the mean service rates were calculated from the data collected and their results were used to measure the performance of the entire system.

Table 4.6: Summary of data collected on Monday

No. of patients	Arrival Time	Service time (minutes)	Time spent in queue(minutes)
1	08:38	00:06	135
2	08:49	00:19	126
3	09:02	00:18	116
4	09:31	00:24	90
5	09:32	00:12	108
6	09:47	00:05	95

4.5.1 Arrival Rates

The number of patients that arrived per unit of time is referred to as mean arrival rate. The counting starts from 8:30 am and ends at 11:30 am.

Calculation of mean arrival rate

The mean arrival rate (λ) is given by

$$\text{Mean arrival rate}, \lambda = \frac{\text{Number of patients in the system between the time interval}}{\text{Number of hours between 8:30 am - 11:30 am}} \text{patients/hr}$$

$$\lambda = \frac{\text{6 patients}}{\text{3 hours}} = \text{2 patients/hour}$$

Results of arrival rates in the week

Table 4.5.1 shows the arrival rates for the week, that is, Monday to Friday. The first column shows the days while the second column shows the arrival rates for each day of the week.

Table 4.7: Arrival Rates in the week

Day	Mean arrival rate (patients/hour)
Monday	2
Tuesday	1.6667
Wednesday	1.6667
Thursday	1.3333
Friday	1.3333

Monday recorded the highest mean arrival rate of 2 patients/hour. The lowest mean arrival rate was 1.3333 patients/hour which was recorded on Thursday as well as Friday. Tuesday and Wednesday also recorded the same value for the mean arrival rate which is 1.6667 patients/hour.

4.5.2 Service Rate

The service rate is the number of patients that are served per unit of time. The mean service rate (μ) is given by,

$$\text{Mean service rate}, \mu = \frac{\text{number of patients}}{\text{total number of hours spent}}$$

Mean service rate for Monday is given by, Total service time = 6+19+18+24+12+5 = 84 minutes = 1.4 hours.

$$\text{Mean service rate}, \mu = \frac{\text{number of patients}}{\text{total number of hours spent}} = \frac{6}{1.4} = 4.2857\,\text{patients/hour}$$

Results of mean service rate for the week

Table 4.5.2 shows the mean service rates for the week. The first column shows the days while the second column shows the mean service rates for the entire week.

Table 4.8: Mean Service Rate for the week

Day	Mean arrival rate (patients/hour)
Monday	4.2857
Tuesday	3.7037
Wednesday	4.7619
Thursday	6.1538
Friday	3.3333

The highest mean service rate was recorded on Thursday with a value of 6.1538 patients/hour whiles the least value of 3.3333 patients/hour was recorded on Friday. Monday and Wednesday recorded mean service rates of 4.2857 patients/hour and 4.7619 patients/hour respectively.

4.5.3 Utilization Factor

The utilization factor (ρ) is the probability that the system is busy when the system is in equilibrium. This is given by,

$$\rho = \frac{\lambda}{\mu}$$

The utilization factor for Monday is given by,

$$\rho = \frac{2}{4.2857} = 0.4667$$

Results of utilization factor for the week

Table 4.5.4 shows the utilization factor for the week. The first column shows the days while the second column shows the utilization factor for each day of the week.

Table 4.9: Utilisation factor for the week

Day	Mean arrival rate (patients/hour)
Monday	0.4667
Tuesday	0.45
Wednesday	0.35
Thursday	0.2167
Friday	0.4

Table 4.5.4 shows that the busiest day of the server was on Monday with 46.67% and followed by Tuesday with a utilization factor of 45%. The least of the utilization factors is 21.67% and occurred on Thursday.

4.5.4 Operating Characteristics

The operating characteristics are:

1. The average time it takes a patient to start being served W_q
2. The average time spent by a patient from arrival until fully served W_s
3. The average number of patients in the system, L_s
4. The average number of patients in the queue, L_q

Calculation of operating characteristics for Monday

1.

$$W_q(\text{average time it takes a patient to start being served}) = \frac{\rho}{\mu - \lambda}$$

$$W_q = \frac{0.4667}{4.2857 - 2} = \frac{0.4667}{2.2857} = 0.2042\,\text{hours}$$

2.

$$W_s(\text{average time it takes a patient from arrival until fully served}) = W_q + \frac{1}{\mu}$$

$$W_s = 0.2042 + \frac{1}{4.2857} = 0.4375\,\text{hours}$$

3.

$$L_s(\text{average number of patients in the system}) = \lambda W_s = 2(0.4375) = 0.875\text{patients}$$

4.

$$L_q(\text{average number of patients in the queue}) = \lambda W_q = 2(0.2042) = 0.4084\text{patients}$$

Results of the operating characteristics for the week

Table 4.5.5 shows the operating characteristics for Monday to Friday. The first column shows the days while the second column shows the operating characteristics for each day of the week.

The highest of the average number of patients in the system was 0.875 patients which was recorded on Monday and the least was recorded on Thursday with 0.2766 patients. Also, the least average number of patients in queue was recorded on Thursday with 0.0599 patients and the highest of 0.4084 patients was

Table 4.10: Operating Characteristics

Day	Operating Characteristics			
	L_s	L_q	W_s	W_q
Monday	0.875	0.4084	0.4375	0.2042
Tuesday	0.8182	0.3682	0.4909	0.2209
Wednesday	0.5385	0.1885	0.3231	0.1131
Thursday	0.2766	0.0599	0.2074	0.0449
Friday	0.6667	0.2667	0.5	0.2

recorded on Monday. The longest waiting time in queue and system are 0.2209 hours and 0.4909 hours respectively and occurred on Tuesday. The shortest waiting time in queue and system are 0.0449 hours and 0.2074 hours respectively and occurred on Thursday.

Table 4.11: Summary of the Operating Characteristics of all the Days

Operating Characteristics	DAYS				
	Monday	Tuesday	Wednesday	Thursday	Friday
μ(patients/hour)	4.2857	3.7037	4.7619	6.1538	3.3333
λ(patients/hour)	2	1.6667	1.6667	1.3333	1.3333
L_s(patients)	0.875	0.8182	0.5385	0.2766	0.6667
L_q(patients)	0.4084	0.3682	0.1885	0.0599	0.2667
W_s(hours)	0.4375	0.4909	0.3231	0.2074	0.5
W_q(hours)	0.2042	0.2209	0.1131	0.449	0.2
ρ	0.4667	0.45	0.35	0.2167	0.4

From Table 4.5.6, the busiest of all the days was Monday since it recorded the highest number of patients and the percentage of time the server is being utilized by a patient (server utilization) was 46.67%. This was followed by Tuesday. Thursday recorded the least of 21.67%. From the table, it also shows that more patients wait in the queue on Mondays than any other day. The average time spent by a patient from arrival until fully served on Monday is 0.4375 hours, which is the highest compared to all the other days. On Thursday, the average time spent by a patient until fully served is 0.2074 hours which is the lowest compared to all the other days. Also, Thursday recorded the lowest average time it takes a patient to start being served whiles Monday recorded the highest followed by

Tuesday, Friday and Wednesday respectively.

4.6 Comparison of Operating Characteristics of the dental department of the Essikado Hospital for different number of servers using average of arrival and service rates.

An average of the arrival and service rates at the dental department are computed by finding the sum of the arrival rates and sum of service rates and dividing by five. The average of the arrival rate during the five days data collection period is given by,

$$\lambda = \frac{2 + 1.6667 + 1.6667 + 1.3333 + 1.3333}{5} = 1.6\text{patients/hour}$$

Also, the average of the service rate during the five days data collection is given by,

$$\mu = \frac{4.2857 + 3.7037 + 4.7619 + 6.1538 + 3.3333}{5} = 4.4477\text{patients/hour}$$

Results of Operating Characteristics for different number of servers

Using the average of the arrival and service rates, the operating characteristics of the system can be calculated for different number of servers as in the table 4.6.1.

Table 4.12: Operating Characteristics

Operating Characteristics for different Number of Servers	Queue Model		
	m/m/1	m/m/2	m/m/3
L_s	0.5619	0.3718	0.3606
L_q	0.2021	0.012	0.0008
W_s	0.3512	0.2324	0.2254
W_q	0.1263	0.0075	0.0005
ρ	0.3597	0.1799	0.1199

From Table 4.6.1 it is observed that there is a reduction in the value of the operating characteristics as the number of servers increases.

An increase in the number of servers from one to two servers indicates that the servers will be 17.99% busy at the dental department. The average number of patients in the system reduced from 0.5619 patients to 0.3718 patients. The average number of patients in the queue decreased from 0.2021 patients to 0.012 patients. The average time spent by a patient from arrival until fully served decreased from 0.3512 hours to 0.2324 hours and the average time it takes a patient to start being served reduced from 0.1263 hours to 0.0075 hours.

Finally, increasing the number of servers from two to three also shows a decrease in the operating characteristics. The server utilisation factor reduced from 17.99% to 11.99%. The average number of patients in the system reduced from 0.3718 patients to 0.3606 patients. The average number of patients in the queue decreased from 0.012 patients to 0.0008 patients. Also, the average time spent by a patient from arrival until fully served reduced from 0.2324 hours to 0.2254 hours and the average time it takes a patient to start being served reduced from 0.0075 hours to 0.0005 hours.

Discussion of Results

The analysis shows that Monday recorded the highest number of patients (six patients) with Thursday and Friday recording the lowest of four patients. The arrival rate of patients on Monday was highest with an arrival rate of 2 patients/hour. The least of the arrival rate was estimated to be 1.3333 patients/hour and occurred on Thursday and Friday. The highest service rate was also found to be 6.1538 patients/hour and occurred on Thursday and the least service rate estimated to be 3.3333 patients/hour occurred on Friday. The least of the average number of patients in the system was estimated to be 0.2766 patients and occurred on Thursday and this corresponded to the least average time spent by a patient from arrival until fully served which is 0.2074 hours. Also, the largest of the average number of patients in the system was estimated to be 0.875 patients and was recorded on Monday whiles the highest of the average time spent by a patient from arrival until fully served was 0.4909 hours recorded on Tuesday.

Using the averages of the arrival and service rates, the operating characteristics decreased with an increase in the number of servers.

Chapter 5

Summary of Conclusions and Recommendations

5.1 Introduction

This chapter summarizes the results of the study, discusses the conclusion arrived at by the researcher and gives recommendations that would be necessary to reduce capacity utilization of the facility and reduce the time a customer spends at the dental clinic.

5.2 Conclusions

Patients' satisfaction is very important to hospital management because the patients are the people who sell the good image of the hospital to others which help to increase the revenue of the hospital. The objective of every health facility is to help reduce patients' waiting time, increase revenue and improve upon customer services and care.

The study looked at the queuing system at the dental department of Essikado Hospital in the Sekondi-Takoradi metropolis. It looked at patients' arrival rates, service rates and the utilization factor of the whole system. These three parameters were then used to measure the waiting time of patients in the queues and in the entire system. They were also used to find the number of patients in the queue and in the whole system.

From the analysis of the study, it was shown that Monday is the busiest day at the dental department because it recorded the highest number of patients (6 patients) with Thursday and Friday recording only four patients. This may be due to the fact that the dental department is closed on weekends so people with dental issues over the weekend have to wait till Monday. During the data collection period, it was observed that most of the patients who spent longer time in the dental treatment room were those who had major dental issues like the extraction of tooth. Patients with minor issues like gum bleeding are those who spent less time at the dental treatment room.

From the analysis, the values of the operating characteristics of the single server model were very high compared to that of the two-server model and three-server model.

5.3 Recommendations

Based on the findings of this study, the following recommendations have been made to help management improve upon patients' satisfaction and also help reduce their waiting times for health care. Patients are uncomfortable with toothache which is very painful and as such will wish to be attended to by a dentist without waiting for long hours. It is therefore recommended that

1. The number of dentist should be increased from one to two in order to reduce the amount of time spent in the queue.

2. Management should obtain a central tray setup system for instruments required for the different dental procedures done in the dental department so as to reduce time spent on sorting them up.

3. It is that the model's waiting time predictions pertain only to waiting time due to server unavailability. The service-end time was recorded as the time patients leave the dental department. I therefore recommend to future researchers to work on the total time a patient spends at the dental department taking into account the time the patient spends at the pharmacy department to take his/her drugs.

4. This research work was done manually and so a simulation model can be developed to confirm the results of the analytical model that was developed

in this work.

REFERENCES

1. Acheampong, L. (2013), Queuing in health care centres, a case studu of out patient department of South Suntreso hospital. Master's thesis, Kwame Nkrumah University of Science and Technology.

2. Adan, I.J.B.F., Boxma1, O.J., Resing, J.A.C. (2000), 'Queuing models with multiple waiting lines,'Department of Mathematics and Computer Science, Eindhoven University of Technology, Eindhoven

3. Adedayo, O. A.; Ojo, O. and Obamiro, J. K. (2006), Operations Research in Decision Analysis and Production Management. Lagos: Pumark Nig. Ltd.

4. Adeleke R. A., Ogunwale O.D., & Halid O.Y. (2009), 'Application of Queuing theory to Waiting Time of Out-Patients in Hospitals', The Paci_c Journal of Science and Technology, Vol. 10, No.-2, November 2009.

5. Agrawal G. and Saxena G. (2010), 'Queuing Model for health care centre' http://74.125.155.132/scholar?q=cache:QsEUUMSV3AQJ: scholar.google.com/& hl=en & as-sdt=2000.

6. Ahmed, S. (2003), 'Accident and Emergency Section Simulation in Hospital' www.wseas.us/e-library/conferences/digest2003/papers/466 124.pdf accessed on 10th December, 2014.

7. Agnihothri, S.R., Taylor P.F. (1991), Sta_ng a Centralized Appointment Scheduling Department in Lourdes Hospital. http:/www.mat.uc.pt/../POMS-Paper.pdf. Accessed 20th January, 2015.

8. Ajay K. S., Rajiv K., Girish K. S. (2013), Queuing theory approach with queuing model: A study. International Journal of Engineering Science Invention, www.ijesi.org Vol. 2, Issue 2, pp 01-24

9. Anokye et al (2013), Application of Queuing Theory to Vehicular Tra_c at signalized intersection in Kumasi Ashanti Region, Ghana. American Journal of Contemporary Research. Vol. 3, No. 7, July 2013.

10. Ashley, D.W., (2000), An introduction to queuing theory in an interactive text format. Transactions on Education. Accessed online from 20/03/15 at http://www.bsbpa.umkc.edu

11. Banks, J., Carson, J. S., Nelson, B. L., Nicol, D. M. (2001), Discrete-Event System Simulation,Prentice Hall international series, 3rd edition, p 24 - 37

12. Bharali, S., (2010), A study on reducing waiting times in the Out-patient Department in a selected hospital. M.Sc Project submitted to Rajiv Gandhi University of Health Sciences.Karnataka, Bangalore.

13. Biggs (2008), 'Hospital waiting list explained', Social Policy Section, accessed online on 4th February, 2015.

14. Biju M. K.; Naeema K. and Faisal U. (2011), Application of Queuing Theory in Human Resource Management in Health Care. ICOQM - 10

15. Bunday, B.D., (1996), An Introduction to Queuing Theory. Halsted Press Publishers, New York: USA.

16. Cayirli and Veral (2003), Outpatient Scheduling in healthcare: A review of literature. Production and Operations Management, Vol. 12, No. 4, pp519-549

17. Crabtree D (2008), Queue (1): Glossary of manufacturing and library of manufacturing topics, 26 glossary pages; Page Q. accessed on 10th December, 2014. Last update April, 2008. Available online at www.glossaryofmanufacturing.com q.html

18. Creemers S. and Lamberecht (2007), 'Queuing Models in Healthcare' *www.econ.kuleuven.be jaargangen* /. . . / TEM-07-3-09-Creemers.pdf accessed on 30th March 2015.

19. Customer Management IQ (2011), Queuing Time: A division of IQPC, IQ Glossary. Available at *http://www.customermanagementiq.com/glossary.cfm*

20. Davis, M.M., Vollman, T.E., (1990), A framework for relating waiting time and customer satisfaction in a service operation. Journal of Services Marketing, 4(1), pp.61-69.

21. Davis, M. M., Acquilano, J.N., (2003), Fundamentals of Operations Management, 4th ed. McGraw Hill Companies Irwin: Boston.

22. David HM (2005), The psychology of waiting line. Q-Matic System company brochure Ashville NC: The Q-Matic Corporation .The McKinsey Quarterly, pp.1-25. Available online at *http://www.qmatic.com*

23. Dhar, Rahman (2013), Case Study for Bank ATM: Queuing model. IOSR Journal of Mathematics, vol. 7, Issue 1 (My-June 2013), pp 01-05

24. Easton and Goodale (2005), Schedule recovery: unplanned absences in service operations, Decision Sciences, 35:459-488

25. Fomundam S. and Herrmann J. (2007), 'A Survey of Queuing theory Applications in Healthcare' drum.lib.umd.edu/bitstream/1903/7222/1/tr-200724.pdf. Accessed 30th March, 2015.

26. Georgievskiy et al (2002), Using Computer simulation modelling to reduce waiting times in emergency department. Accessed at http://www.flexism.com /products/healthcare/dics/Reduce_ER_wait_times.pdf on 5th March, 2015.

27. Fink and Gillett (2006), Queuing theory and the Taguchi Loss Function: The cost of customer dissatisfaction in waiting lies. International Journal of Strategic Cost Management/Spring 2006

28. Foster et al (2010), 'A Spoonful of Math helps the medicine Go Down : An Illustration of How Healthcare benefit from mathematical modeling and

analysis', BMC Medical Research

29. Methodology 2010, 10:60 http://www.biomedcentral.com/1471-2288/10/60. Accessed 2nd March, 2015.

30. Gorunescu et al (2002, A queue model for bed-occupancy management and planning of hospitals. Journal of the Operational Research Society. 53, 19-24

31. Gorney, L., (1981), Queuing Theory. A Problem Solving Approach. Petrocelli Books Inc: New York.

32. Green, L.V., Soares, J., Green, R.A., (2006), Using Queuing Theory to Increase the Effectiveness of Emergency Department Provider Staffing. Accessed from Acad Emergency Med., Jan; 13(1):61 - 8.Epub in October (2012)

33. Hall et al (2006), Discrete-event simulation of health care systems, in Patient Flow: Reducing Delay in Healthcare Delivery, Hall, and R.W. Ed. Springer: New York.

34. Hartvigsen (2004), 'SimQuick Process Simulation with Excel, Second Edition'. Accessed at dev.prenhall.com/divisions/bp/app/.../SimQuick/SimQuick.../SimQ on 30th January, 2015.

35. Hiller, F. and Lieberman, G. (1974), Introduction to Operations Research: Sixth Edition. New York: McGraw-Hill, Inc.

36. Hiller, F. and Lieberman, G. (2001), Introduction to Operations Research, McGraw-Hill Higher Education, 7th Edition, p834-8

37. Imran et al (2014), Measuring Queuing System and Time Standards: A case study of student a_airs in universities. Academic Journal of Business Management vol 8(2) pp 80-88

38. Katz, K., Larson, K., Larson, R., (1991).Prescription for the waiting in - line - Blues: Entertain, Enlighten and Engage', accessed in November 2014 at www.zotero.org.

39. Katz, K.L., Martin, B.R., (1989, Improving Customer Satisfaction through the Management of Perceptions of Waiting. M.Sc Project submitted to the Sloan Management, MIT.

40. Kano, N., (1984), Attractive Quality and Must-Be Quality. Journal of the Japanese Society for Quality Control (JSQC), Volume 14, pp. 145-146

41. Kembe et al (2012), A Study of Waiting And Service Costs of a Multi-Server Queuing Model in a Specialist Hospital. International Journal of Scientific and Technology Research, vol. 1, Issue 8.

42. Klinrock, L. (1976), 'Queueing Systems Vol. II: Computer Applications', John Wiley & Sons, New York.

43. Kolker (2008), Queuing Analytic Theory and Discrete Events Simulation for Healthcare. www.iienet.org/.../Queuing%20Analytic%20Theory20% 20and%20Discrete%

44. Kolmogorov and Feller (1958), An Introduction to Probability theory and its applications. American Math Society, Vol. 1, 2nd edition, 64(6):393

45. Koizumi. N., Kuno, E., Smith, T.E., (2005), Modelling Patient Flows Using a Queuing Network with Blocking. Health Care Management Science, 8, 49-60.

46. Kothari, C.R., (2009), Quantitative Techniques, 3rd ed. Vikas Publishing House PVT ltd. New Delhi.

47. Kotler P. (1999), Marketing Management, The New Millennium Edition

48. Kuran S., Karunesh S., (2011). A study of applicability of waiting line model in healthcare: A systematic review. JMT, volume 19, Number 1. January-June 2011.

49. McClain (1976), Bed Planning using Queuing Theory models of hospital occupancy: a sensitivity analysis. Inquiry, 13, 167-176

50. Mackey and Cole (1997), Patient Waiting time in Nursing Managed Clinic. The International Journal Adv. Nur. Practice, 1 p1

51. McQuarri, D. G. (1983), "Hospital Utilization levels. The Application of Queuing Theory to a controversial Medical Economic Problem. Minnesota

Medicine, 66, 679-686

52. McManus, M.L., Long, M.C., Cooper, A., Litvak, E., (2004), Queuing Theory Accurately Models the Need for Critical Care Resources. Anesthesiology, 100, 1271-1276.

53. Medhi (2003,) Stochastic models I Queuing theory. 2nd Edition, Academic Press, California

54. Mehandiratta R. (2011), 'Application of Queuing theory in Health Care' International Journal of Computing and Business Research, ISSN (online): pp 2229 - 6166, Vol. 2 Issue 2 .

55. Mohammed et al (2014), Using Queuing Theory and Simulation to Optimise Hospital Pharmacy Performance. Iranian Red Crescent Medical Journal, 16(3)

56. Molina, E.C., (1927), Application of the Theory of Probability to Telephone Trucking Problems. www.lit.upcomillas.es/timeline.pdf. Accessed 12th February, 2015.

57. Moss (1987), Hospital Pharmacy Staffing levels and outpatient waiting times. Pharm J, 239:69-70

58. Ndukwe H. C., Omale S. and Opanuga O. O.(2011), Reducing queue in a Nigerian hospital pharmacy. African Journal of Pharmacy and Pharmacol-

ogy Vol. 5(8). Pp 1020-1026.

59. Nkeiruka et al (2013), Application of Queuing Theory to patient satisfaction at a tertiary hospital in Nigeria. Nigerian Medical Journal, 54(1): 64-67

60. Nosek, A.R., Wilson, P. J., (2001), Queuing Theory and Customer Satisfaction: A Review of Terminology, Trends and Applications to Pharmacy Practice. Hospital Pharmacy, 36(3),pp. 275-279.

61. Nuti S, Vainieri M (2012), Managing waiting times in diagnostic medical imaging. BMJ Open 2012; 2:001255. Doi:1.1136/bmjopen 2012-001255

62. Obamiro, J. K. (2010), Queuing Theory and Patient Satisfaction: An Overview of Terminology and Application in Ante-Natal Care Unit. Petroleum-Gas University of Ploiesti Bulletin, Economic Sciences Series, Vol. LXII No. 1/1 - 11. http://www.upgbulletin-se.ro 2/8/2014. Accessed on 2nd May, 2015

63. Parasuraman, A., Zeithamil, V.A., Berry, L.L., (1985), A Conceptual Model of Service Quality and its Implications for Future Research, Journal of Marketing. Volume 49, pp. 41-50.

64. Parasuraman, A., Zeithamil, V.A., Berry, L.L., (2009), Delivering Quality Service: Balancing Customer Perceptions and Expectations. 13th ed. New York: Free Press.

65. Pollaczek (1965), "Concerning an analytical method for the treatment of queuing problems". Proc. Symp. Congestion theory, University of North Carolina, Chapel Hill, 1-42

66. Preater, J. (2002), Queues in Health. Health Care Management Science, 5, 283

67. Obamiro, J. K., (2006), Operations Research in Decision Analysis and Production Management. Pumark Nig. Ltd., Lagos.

68. Sanish (2007), Application of Queuing model and Simulation to the traffic at New Mangalore Port Department of Applied Mechanics and Hydraulics, NITK Surathkal, Karnataka 69. Schoenmeyr et al (2009), A Model for understanding the Impacts of Demand and Capacity on waiting time to enter a Congested Recovery Room. Anesthesiology, Vol. 10, No.6

69. Singh, V., (2006), Use of Queuing Models in Health Care, Department of Health Policy and Management, University of Arkansas for medical science. Accesed in February 2015 at www.u amont.edu.

70. Shimshak, D.G., Gropp Damico, D., Burden, H.D., (1981), A priority Queuing Model of a Hospital Pharmacy Unit. European Journal of Operational Research, 7, 350-354.

71. Siddhartan, K., Jones, W.J., Johnson, J.A., (1996), A Priority Queuing Model to Reduce Waiting times in Emergency Care. International Journal

of Health Care Quality Assurance, page 10-16.

72. Sundarapandian (2009), Probability, Statistics and Queuing Theory. PHI Learning Pvt. Ltd., New Delhi.

73. Stakutis C, Boyle T (2009), Your health, your way: Human-enabled health care. CA Emerging Technologies, pp. 1-10.

74. Vandaele et al (2003), Optimal Grouping for a nuclear magnetic resonance scanner by means of an open Queuing model, European Journal of Operations Research, 151, 181-192

75. Vohhra, N.D., (2010), Quantitative Techniques in Management. 4th ed. Tata McGraw Hill Education Private Ltd, New Delhi.

76. Woensel and Cruz (2009). A Stochastic Approach to Traffic Congestion Costs, Computers and Operations Research, 36(6), 1731-1739

77. Winston and Albright (1997), Practical Management Science: Spread sheet Modelling and Applications. Belmont, CA: Duxbury.

78. Yariv et al (2012), IIE Transactions on Healthcare Systems Engineering (2012)2, 233-247.

79. Yeung, M. W. S., Herrison, P. G., and Knottenbelt, W. J. (2006) " A Queuing Network Modal of Patient Flow in an Accident and Emergency

Department" Unpublished Paper.

80. Zhang et al (2000), Simulation in Queuing Model using simulation at Beiteba crossing check point. Sixth Youth Science Conference, Ministry of Education, Singapore.

REFERENCES

Baidoo, A. (2013). The spread of hiv/aids in ghana. *International journal for the Spread HIV/AIDS*, 35:210–240.

Barnett, W. A., He, Y., and Yansi, F. (2002). Stabilization policy as bifurcation selection: Would stabilization policy work if the economy really were unstable. ,. *Journal of Science Maths*, 5:1025–1052.

Okyere, G. A. (2013a). *Introduction to LaTex*.

Okyere, G. A. (2013b). Mathematics is for great thinkers. *Ghana Mathematical Journal*, 34:23.