KWAME NKRUMAH UNIVERSITY OF SCIENCE AND

TECHNOLOGY

AN APPLICATION OF QUEUING THEORY TO OUTPATIENTS WAITING TIME;

A CASE STUDY OF THE SEVENTH DAY ADVENTIST HOSPITAL TAMALE

By

ISSAKA H. ALHASSAN

(PG4066910)

A Thesis submitted to the Department of Mathematics,

Kwame Nkrumah University of Science and Technology

In Partial Fulfilment of the Requirements for the Degree of Msc. Industrial

Mathematics Department of Mathematics

Institute of Distance Learning

November 2012

DECLARATION

I hereby declare that this submission is my own work towards the Master of Science Industrial Mathematics, that to the best of my knowledge. It contains no material previously published by another person or material which has been accepted for the award of any other degree of the University. Related works by others, which served as source of information, has been duly acknowledged.

Issaka H. Alhassan .... 08/01/13 ....
(PG4066910) Signature Date

Certified by:

Mr. Charles Sebil .... 08/01/13 ....
Supervisor Signature Date

Certified by:

Mr. F. K. Darkwah ....
HOD Dept. of Mathematics Signature Date

ABSTRACT

This thesis considers the average waiting time of patients at the Out-patient department (OPD) of Tamale Seventh Day Adventist hospital as a single-channel queuing system with Poisson arrivals and exponential service rate where arrivals are handled on a first come first serve (FCFS) basis. Hence, the m/m/1 queuing system is however proposed.

The total waiting time in queues at the four sections of the OPD before service is completely rendered is 2.67 hours (160.2 minutes) and the average number of patients waiting for service at any given hour is 65. The server utilization at the Records/health insurance section, History section, consulting rooms and the dispensary is 88%, 94%, 91%and 88% respectively. The probability of a patient queuing at each of the section at the OPD is 0.88, 0.94, 0.91 and 0.88 respectively.

The three main causes identified as the contributing factors towards excessive patient's waiting time in this thesis are: the registration time, insufficient number of counter service staff and insufficient number of doctors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

DEDICATION

This thesis is dedicated to my parents Mallam Issaka and Mum Zenabu, my beloved wife Faridatu, my son Sibawiala Ramin and to all who have contributed in one way or the other to my success in life.

AKNOWLEDGMENT

Many thanks to the almighty God, the provider of knowledge and wisdom for his guidance throughout this program, my life and the write up of this thesis.

I am highly indebted to my supervisor, Mr Charles Sebil for his guidance throughout this research work. God richly bless you.

To my wife Faridatu, thank you for all the encouragement and support you have given me all along.

Finally to my Mother Zenabu and my siblings, I say, thank you for everything you have done for me.

CHAPTER ONE

1.0 INTRODUCTION

1.1 Background

The Seventh Day Adventist (SDA) Hospital is one of the leading hospitals in the Tamale Metropolis of the Northern regions of Ghana. This hospital provides a number of patient services and it is mainly responsible for delivering a comprehensive range of primary and secondary care.

The outpatient department is a vital component of the health care system. The Out Patient Department of the Tamale Seventh Day Adventist Hospital comprises of General OPD section, Admission section and Emergency Treatment Unit (ETU). This study is focused on only the General OPD section.

The continuous increase in number of patients at the OPD leads to the long waiting time of patients in the hospitals. This problem is a serious issue in the health care system. Waiting time in outpatient departments is a major concern throughout the world. One consistent feature of patient dissatisfaction has been expressed with the length of waiting time in the outpatient department. The waiting time is particularly important for any hospital, since the "customers' are 'patients". Long waits create customer dissatisfaction on one hand and resources inefficiencies on the other hand.

In Ghana, patients also spend long waiting time in hospital especially in the public and private hospital where the National Health Insurance Scheme is applicable. This long waiting time leads to patients' dissatisfaction and some patients leave without getting any services. For these reasons, hospitals are making efforts to improve their service system to increase patient satisfaction.

The outpatient department represents a complex system through which many patients with varying needs pass each day. An effective appointment system is a critical component in controlling patient waiting times within clinic sessions. Current waiting times are often unacceptable and place great stress on clinic staff. There are two scheduling systems in outpatient departments: the patient and staff. Appointment scheduling and staff scheduling are the two aspects that determine the waiting time in the outpatient departments.

A proper service system should be designed in order to reduce waiting time of patients in the system and increase patients' satisfaction.

Findings from the research could be used by hospital administrators to address gaps in human resources, logistics, infrastructure and other internal procedures towards ensuring an effective and efficient health care delivery system at the Hospital.

1.2 Statement of the Problem

Access to health care remains a defining element in Ghana. However as the demographic and epidemiological characteristics of the population change, timely access is becoming more elusive and difficult to ensure. Multiple demands on the current health care system are being felt across the spectrum of health care service delivery with much attention toward wait times for health services.

The introduction of the National Health Insurance Scheme (NHIS) into the Ghanaian health care system has made the issue of timely access to health care more difficult and creating long queues at the various health centres.

The Seventh Day Adventist Hospital being centrally located in the metropolis and closer to the commercial centre receives all categories of patients across the region. Even though there are other health facilities within Tamale and its suburbs that accept the National Health Insurance Scheme (NHIS), a lot of patients within Tamale prefer to seek medical care at the SDA hospital. This situation leads to long queues and congestion at the outpatient department. In some clinic sessions, the total number of patients and accompanying relatives/friends far exceeds the number of available seats in the outpatient unit.

One consistent feature of patient dissatisfaction has been expressed with the length of waiting time in the outpatient department before being served. Some patients end up spending the whole day at the outpatient department of the hospital. This long waits and queues create customer dissatisfaction on one hand and resources inefficiencies on the other hand.

A study of the out-patients waiting time at the Seventh Day Adventist Hospital is critical to public appreciation of the quality of health care operating environment at the hospital; hence, this study is aimed at assessing patients' waiting time and factors affecting waiting in the outpatients' departments.

1.3 Objectives

The objectives of this thesis are to:

1. To model the service to patients at the outpatient department as queuing theory.
2. To solve the model by First Come First Serve (FCFS) method of queuing theory to determine the average waiting time for service, resource utilization at the hospital and to suggest ways to optimize waiting time at the outpatient department

1.4 Data Collection

Data on outpatient daily attendance at the Seventh Day Adventist Hospital was both primary collected by personnel using stop watches and secondary collected from the record department of the Hospital. The outpatient department comprises of: The records unit/ National Health Insurance unit, History unit, Consultation unit, Dispensary unit and sometimes Laboratory unit.

1.5 Scope of the Study

The study aims at studying the pattern of queues and to analyse the current situation of waiting time and overcrowding at the out-patients department at the Seventh Day Adventist Hospital and to suggest ways to provide efficient service to patients.

The study is limited to the outpatient department. This implies that only patients who end their treatment at the outpatients unit are considered in this study.

Additionally, this thesis takes into consideration the estimated waiting time but ignores the effect of perception of waiting time by patient and their relatives.

1.6 Organization of the work

This thesis is organized into five chapters. Chapter one constituted the introduction, which focuses mainly on the background, statement of the problem, objectives, the scope and limitation and significance of the study. Review of the theoretical and empirical literature pertinent to the concern of the thesis is presented in Chapter two. Chapter three describes the research methodology that includes a brief description of the study area, data collection procedures and analytical techniques. The chapter four looks at the results of the study and discussions. Finally, the summary of the major findings, conclusion and recommendation are presented in Chapter five.

CHAPTER TWO

2.0 LITERATURE REVIEW

2.1 Introduction

The organizations that care for persons who are ill and injured vary widely in scope and scale, from specialized outpatient clinics to large, urban hospitals to regional healthcare systems. Despite these differences, one can view the healthcare processes that these organizations provide as queuing systems in which patients arrive, wait for service, obtain service, and then depart. The healthcare processes also vary in complexity and scope, but they all consist of a set of activities and procedures (both medical and non-medical) that the patient must undergo in order to receive the needed treatment. The resources (or servers) in these queuing systems are the trained personnel and specialized equipment that these activities and procedures require.

This chapter is divided into two parts. The first part looks at literature on queuing in general and the second part looks at related literature in queuing in hospitals.

2.1 Literature on Queuing

A considerable body of research has shown that queuing theory can be useful in real-world healthcare situations, and some reviews of this work have appeared.

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 dispensary 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 (2006a) presents the theory of queuing as applied in healthcare. She discusses the relationship amongst delays, utilization and the number of servers; the basic M/M/s model, its assumptions and extensions; and the applications of the theory to determine the required number of servers.

2.1.1 The Psychology of Queuing and the Effect of the Number of People Front and Rear

Psychologist Festinger (1954) brought up social comparison theory, which suggested that people have the needs or drivers to assess their own abilities and opinions). Hereby, when people are not sure about personal abilities and opinions and are lack of objectively assessing standards, they will try to figure self' uncertainness via comparing to similar others.

Gross & Harris (1985) reported that operation researchers have widely studied the structure and management of queues with the objective of developing efficient queuing policies from the perspective of the entire system. Much of this research is based on mathematical modelling, but it typically ignores consumer experiences in queue.

Larson (1987) reported that waiting for service is typically a negative consumer experience and cause unhappiness, frustration, and anxiety.

Katz et al. (1991) found that distraction during the waiting period (e.g., a news board or television) made the wait more joyful and improved service evaluation.

Meyer (1994) argues that consumers are not mindless passengers in a human line but can make decisions and their mood states are also open to environmental influence.

A large number of people rear the consumer imply a longer queue, ceteris paribus. A consumer may make two inferences on seeing a longer queue: (1) it may serve as a social validation cue that the service is worth waiting for Cialdini (1985), and (2) it may lead the consumer to expect a longer queue if she re-join at a later point in time. Both these inferences will result in a greater unwilling to leave the queue. Specifically, while waiting is unpleasant, seeing many people rear is somewhat of a comfort because there are people worse off than me Zhou & Soman (2003).

Cialdini, (1985) Researches in social psychology focus on the universal human tendency to learn about and improve that by making social comparisons with others. Social comparisons occur on an ongoing basis, and these have been described as spontaneous, effortless, and relatively automatic (Gibbons & Buunk 1999). Therefore, one reason the number of people rear matters may be related to social comparisons, more specifically, downward comparisons directed toward self-enhancement (Wills, 1981). Another reason the number of people rear matters may be related to social facilitation. Social facilitation assumes that the presence of other people in the setting will result in positive affect because others fill the time in distracting and entertaining ways (Baker & Cameron, 1996).

Maister (1985) argued that promoting group waiting will increase customers' tolerance for waiting time because it distracts people from paying attention to the time. The concept of social facilitation may explain that people waiting in line sometimes will interact by discussing the wait, sharing frustration, and consoling each other (Baker & Cameron, 1996).

In the psychology of queuing situation, people will exercise upward (to those who are in front) and downward (to those who are behind) comparisons. Nevertheless, basing on previous literature, while having comparison downward, consumers are more likely to have anxious and unhappy feelings for their queuing position (Wills, 1981). This is due to the negative consuming experiences by waiting for services. It will cause the uncertainty, unhappiness, bore, frustration, nervous, irritation, and anxiousness to consumers (Carmon, Shanthikumar, & Carmon, 1995; Katz et al., 1991; Larson, 1987, Leclerc et al., 1995; Maister, 1985; Zhou & Soman, 2003). Consumers who are queuing will try to ease their anxiousness and irritation by comparing to those who are more unfortunate behind them.

(Hui & Tse, 1996; Zhou & Soman, 2003) stated that as the time of waiting by the consumer gets longer, yet the people count waiting in front also gets higher, then negative mood will be stronger, service evaluation naturally will be lower. This kind of phenomenon is the psychology of waiting in line called "The Number Ahead Effect". On the other hand, according to "The Number Rear Effect" (Zhou & Soman, 2003), the more people queuing behind you, the stronger positive feelings you get.

2.1.2 Perceived Waiting Time

Because services cannot be inventoried and demand may be hard to predict, waiting before receiving a service is a common experience.

Bitner (1992) proposed that the service environment can affect consumers' emotional, cognitive, and physiological responses, which in turn influence their evaluations and behaviours. Some scholars extend Bitner's work by exploring the effects of specific environmental elements on one intervening variable—perception of waiting time—that has been found to influence service evaluation (Clemmer & Scheeider, 1989). Perception of waiting time may influence mood states and thus influence overall evaluation of the service encounter (Baker & Cameron, 1996).

Leclerc et al. (1995) examine whether waiting-time decisions are in fact similar to monetary decisions. In actual life, when we order Domino's pizza, it will give us one hundred dollars coupon if it doesn't deliver the pizza in thirty minutes. Offering compensation to customers in the case of services failure may give sales people some ability to count this effect (Widmier & Jackson, 2002). Research in the area of service failure gives some indication that compensation for a service failure does have positive impact on customer satisfaction with services encounter.

According to past literature, the effect of the number of people front or rear on the mood states will be moderated by the salience of the relative position as determined by the queue system; that is, the more salient the relative position, the greater the number front or rear effect (Zhou & Soman, 2003). In the queuing context, the information about the relative position of the consumer will therefore influence the perception of waiting time. A linear queue system, in which people queue up one behind another in order of arrival, offers clear information about the position of a consumer relative to others in the waiting-area. A review of the literature suggests that perception of waiting time moderates the influence of mood states via four routes: (1) directly, (2) indirectly through queuing perceptions, (3) indirectly through attribution of the delay, or (4) indirectly through degree of filled time(i.e., the cognitive timer) (Clemmer & Scheeider, 1989). Individuals who find the waiting time unacceptable have a very significantly lower mood and perceived the service as being of lower quality.

2.1.3 The Mood States of the Consumer

Gardner, (1985) explained that the word "mood" has a wide range of usages and meaning. One might use the term to describe a phenomenological property of an individual's subjectively perceived affective state; e.g., someone may be in a cheerful mood or a hostile mood. One might also use mood to describe a property of an inanimate object; e.g., a point-of –purchase display may have a "sophisticated mood" or a "fun mood".

Belk (1984) concluded that we adopt the former, phenomenological, approach and view moods as feeling states that are subjectively perceived by individuals. Although categorizing moods as positive or negative may be an oversimplification, different types of positive moods (e.g., cheeriness, and peacefulness) and negative moods (e.g., anxiety, guilt, and depression) can be readily identified. The states of the mood contain positive or negative parts. Positive mood means peacefulness, satisfaction, relaxation, or a kind of happy feeling. Negative mood refers to depression, anxiety, or anger. The literatures of many functions for inquiring the moods point out that both of them will affect the process of the social behaviour, thinking and judgment of people (Isen, 1987).

2.1.4 Waiting in Queues

Literature on queuing indicates that waiting in lines or queues causes inconvenience to economic costs to individuals and organizations. Hospitals, airline companies, banks, manufacturing firms etc., try to minimize the total waiting cost, and the cost of providing service to their customers. Therefore, speed of service is increasingly becoming a very important competitive parameter (Katz, et al, 1991).

Davis et al (2003) assert that providing ever-faster service, with the ultimate goal of having zero customer waiting time, has recently received managerial attention for several reasons. First, in the more highly developed countries, where standards of living are high, time becomes more valuable as a commodity and consequently, customers are less willing to wait for service.

Second, this is a growing realization by organizations that the way they treat their customers today significantly impact on whether or not they will remain loyal customers tomorrow.

Finally, advances in technology such as computers, internet etc., have provided firms with the ability to provide faster services. For these reasons hospital administrators, physicians and managers are continuously finding means to deliver faster services, believing that the waiting will affect after service evaluation negatively. Also, understanding the inefficiencies in the hospital and improving them is crucial for making health care policy and budgeting decisions (Wilson and Nguyen, 2004).

Cochran and Bhati (2006) also argue that higher operational efficiency of the hospital is likely to help to control the cost of medical services and consequently to provide more affordable care and improve access to the public. Addressing the problems of queuing involve a trade-off between the costs of customers waiting time and the cost of providing faster service.

Katz et al (1991) reported that researchers have argued that service waits can be controlled by two techniques: operations management or perceptions management. The operation management aspect deals with the management of how patients (customers), queues and servers can be coordinated towards the goal of rendering effective service at the least cost.

Sitzia and Wood (1997) stated that Patients' perception of health care has gained increasing attention over the past 20 years. Patients' 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 amount of time customers must spend waiting can significantly influence their satisfaction (Davis and Vollman, 1990).

Taylor (1994) reported that research has demonstrated that customer satisfaction is affected not just by waiting time but also by customer expectations or attribution of the causes for the waiting.

Consequently, one of the issues in queue management is not only the actual amount of time the customer has to wait, but also the customer's perceptions of that wait (Davis and Heineke, 1994).

Obviously, there are two approaches to increasing customer satisfaction with regard to waiting time: through decreasing actual waiting time, as well as through enhancing customer's waiting experience (Katz, Larson, and Larson, 1991; Davis and Heineke, 1994).

Nosek and Wilson (2001) defined Queuing theory as basically a mathematical approach applied to the analysis of waiting lines within the field of operations management.

(Singh, 2006) stated that any system in which arrivals place demand upon a finite capacity resource may be termed as a queuing system. In the case of ante-natal care unit, it can be found that pregnant women arrive or demand services randomly. The objective of queuing analysis and its application in health organizations is to "minimize costs" to the organization- both tangible and intangible. The rising cost of health care can be attributed not only to ageing population expensive and advanced treatment modalities but also to inefficiencies in health delivery. Queuing theory application is an attempt to minimize the cost of providing health care services through minimization of inefficiencies and delays in the system.

Gorney, 1981; Bunday, 1996 explained that Queuing theory uses queuing models or mathematical models and performance measures to assess and hopefully improve the flow of customers through a queuing system. A good patient flow means that the patient queuing is minimized while a poor patient flow means patients suffer considerable queuing delays (Hall, 1999).

Queuing theory has many applications and has been used extensively by the service industries (Nosek and Wilson, 2001).

A queuing system or waiting line phenomenon consist essentially of six major components: the population, the arrival, queues itself, queue discipline, service mechanism, departure or exit.

The population source serves as where arrivals are generated. Arrivals of patients at the hospital may be drawn from either a finite or an infinite population. A finite population source refers to the limited size of the customer pool. Alternatively, an infinite source is forever.

Taha (2005) wrote that queue discipline is the sequence in which customers or patients are processed or served. The most common discipline is first come, first served (FCFS). Other disciplines include last come, first served (LCFS) and service in random order (SIRO). Customers may also be selected from the queue based on some order of priority.

The service mechanism describes how the customer is served. It includes the number of servers and the duration of the service time-both of which may vary greatly and in a random fashion (Nosek and Wilson, 2001). The number of lines and servers determines the choice of service facility structures. The common service facility structures are: single-channel, single–phase; single-channel, multiphase; multi-channel, single phase and multi-channel, multiphase.

The departure or exit occurs when a customer is served. The two possible exit scenarios as mentioned by Davis (2003) are: (a) the customer may return to the source population and immediately become a competing candidate for service again; (b) there may be a low probability of re-service.

2.2 Application of Queuing Theory in Health Care

(Bailey 1952) noted that the application of queuing theory within the healthcare sector has been in existence for over fifty years. It was noted by O'Keefe (1985) as being one of the early examples of the use of operational research.

Bailey(1952) in his paper developed a mathematical model of the outpatient queuing process for a general practitioner's clinic, commencing with the appointment, moving through the arrival and on to the patients being seen by the doctor, a model which could still be applied to today's outpatient clinic in most hospitals. Little, it would seem has changed. At least that is probably what the patients believe, for there are still queues, both real and virtual. Long endless waits for an appointment to be confirmed possibly followed by a cancellation and new dates, and finally waiting for the day to arrive (the virtual queue). And then eventually on the day, unbelievable and unexplained waits before the short but valued consultation. Bailey (1952) found that the majority of patients arrive on time for their appointments

Singh (2006) reported that the health systems ability to deliver safe, efficient and smooth services to the patients did not receive much attention until mid-1990's. Several key reimbursement changes, increasing critiques and cost pressure on the system and increasing demand of quality and efficacy from highly aware and educated patients due to advances in technology and telecommunications, have started putting more pressure on the healthcare managers to respond to these concerns.

Hall, (1991) noted that queuing theory manages patient flow through the system. If patient flow is good, patients flow like a river, meaning that each stage is completed with minimal delay. When the system is broken, patients accumulate like a reservoir. Healthcare systems resemble any complex queuing network in that delay can be reduced through: (1) Synchronization of work among service stages, (ii). Scheduling of resources (e.g. doctors and nurses) to match patterns of arrival and, (iii) constant system monitoring (e.g. treating number of patients waiting by location, diagnostic grouping) linked to immediate actions. Recently, application of stochastic methods has increased in analysing clinical problems (Kandemir-Cavas and Cavas, 2007). Queuing theory, as the most common application of the stochastic process, examines queues or waiting lines dealing with random input and servicing processes (Wu, 1998).

2.2.1 Reneging

When a patient is waiting in a queue, he may decide to forgo the service because he does not wish to wait any longer. This phenomenon, called reneging, is an important characteristic of many healthcare systems. 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). An important example of such a system is an emergency department.

Broyles and Cochran (2007) calculated the percentage of patients who leave an emergency department without getting help using arrival rate, service rate, utilization, capacity. From this percentage, they determine the resulting revenue loss.

It is possible to redesign a queuing system to reduce reneging. A common approach is to separate patients by the type of service required. Roche et al. (2007) find that the number of patients who leave an emergency department without being served is reduced by separating non-acute patients and treating them in dedicated fast-track areas. Most of their waiting would be for tests or test results after having first seen a doctor. The paper also estimates the size of the waiting area for patients and those accompanying them.

2.2.2 Variable Arrival Rate

Although most analytical queuing models assume a constant customer arrival rate, many healthcare systems have a variable arrival rate. In some cases, the arrival rate may depend upon time but be independent of the system state. For instance, arrival rates change due to the time of day, the day of the week, or the season of the year. In other cases, the arrival rate depends upon the state of the system. A system with congestion discourages arrivals.

Worthington (1991) suggests that increasing service capacity (the traditional method of attempting to reduce long queues) has little effect on queue length because as soon as patients realize that waiting times would reduce, the arrival rate increases, which increases the queue again. Worthington (1987) presents an $M(\lambda q)/G/S$ model for service times of any fixed probability distribution and for arrival rates that decrease linearly with the queue length and the expected waiting time. The arrival rate may increase over time due to population growth or other factors. Rosenquist (1987) studies how an increase in patient arrival rate affects waiting times and queue length for an emergency radiology service.

2.2.3 Priority Queuing Discipline

In most healthcare settings, unless an appointment system is in place, the queue discipline is either first-in-first-out or a set of patient classes that have different priorities (as in an emergency department, which treats patients with life-threatening injuries before others).

McQuarrie (1983) shows 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.

When arriving patients are placed in different queues, each of which has a different service priority, the queue discipline may be pre-emptive or non-pre-emptive. In the latter, low priority patients receive service only when no high priority patients are waiting, but the low priority patient who is receiving service is not interrupted if a high priority patient arrives and all servers are busy. In the pre-emptive queue discipline, however, the service to a low priority patient is interrupted in this event. Green (2006a) presents models for both queue disciplines.

Siddhartan et al. (1996) analyse the effect on patient waiting times when primary care patients use the Emergency Department. They propose a priority discipline for different categories of patients and then a first-in-first-out discipline for each category. They find 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.

Haussmann (1970) investigates the relationship between the composition of prioritized queues and the number of nurses responding to inpatient demands. The research finds that slight increases in the number of patients assigned to a nurse and/or a patient mix with more high-priority demands result in very large waiting times for low priority patients.

Worthington (1991) analyses patient transfer from outpatient physicians to inpatient physicians. The patient is assigned one of three priority levels. Based on the priority level, there is a standard time period before which a referred patient should be scheduled to see the inpatient physician. The model assumes sufficient in-patient capacity to treat the highest priority category within its standard time, and proposes sharing the remaining service capacity amongst the lower priority levels in such a manner that they each exceed their standard target times by the same percentage.

Taylor et al. (1969) modelled an emergency anaesthetic department operating with priority queuing discipline. They are interested in the probability that a patient would have to wait more than a certain amount of time to be served.

Fiems et al. (2007) investigated the effect of emergency requests on the waiting times of scheduled patients with deterministic processing times. It is a pre-emptive repeat priority queuing system in which the emergency patients interrupt the scheduled patients and the latter's service is restarted as opposed to being resumed. This paper models a single server queue and divides time into equally long slots (discretizing time). Periods of emergency interruptions are considered to have no server available from the point of view of the scheduled patients (vacation). The result is a discrete-time queuing model with exhaustive vacations.

2.2.4 Blocking

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) present 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 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.

2.2.5 System Design

Because patient waiting is undesirable, limiting waiting times is an important objective when designing a healthcare system. This section reviews work on determining system capacity based on desired system goals and requirements. The variables of interest are usually staffing levels, beds, or other key resources.

Bailey (1954) first establishes the existence in outpatient and inpatient clinics of a threshold capacity which occurs at the point where service supply equals demand. When the number of servers is below this threshold, a clinic develops an infinite queue. Slightly above this threshold, waiting time and queue length are low. He argues that it is therefore sufficient to design for a capacity that exceeds the expected demand (with stochastic error accounted for) by a value of 1 or 2. Long waiting lists are most likely the result of accumulated backlog which can be depleted by a temporary surge in supply. Seasonal variations in supply would also result in a sharp rise in waiting list length.

Moore (1977) reduces customer waiting time for birth and death certificates at the Dallas bureau of vital statistics by decreasing the time required to serve each customer. This research first uses queuing theory to calculate the service rate required to achieve a target waiting time of 15 minutes. This service rate is converted to the time required to serve one customer. The reduced time required to serve each customer is attained through the use of new equipment and more efficient processes.

Agnihothri and Taylor (1991) seek the optimal staffing at 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 paper grouped periods that receive similar call intensity and determines the necessary staffing for each such intensity, so that staffing varies dynamically with call intensity. As a result of redistributing server capacity over time, customer complaints immediately reduced without an addition of staff. Green (2006b) uses the same approach and names it Stationary Independent Period by Period (SIPP) to adjust staffing in order to reduce the percentage of patients that renege. However, arguing that congestion starts sometime after the arrival peak, the staffing levels should lag behind the service demand levels (lag SIPP).

2.2.6 Minimize Costs

Determining server capacity by minimizing the costs in a healthcare queuing system is a special case of system design. Most of the research assigns costs to patient waiting time and to each server. After modelling the system using queuing theory, minimizing costs reduces to an exercise of finding the resource allocation that costs the least or generates the most profit.

Keller and Laughhunn (1973) set out to determine the capacity with minimal costs required to serve patients at the Duke University Medical centre. They find that the current capacity is good but needs to be redistributed in time to accommodate patient arrival patterns.

Young (1962a, b) proposes an incremental analysis approach in which the cost of an additional bed is compared with the benefits it generates. Beds are added until the increased cost equals the benefits.

Shimshak et al. (1981) consider a dispensary queuing system with pre-emptive service priority discipline where the arrival of a prescription order suspends the processing of lower priority prescriptions. Different costs are assigned to wait-times for prescriptions of different priorities.

Gupta et al. (1971) choose the number of messengers required to transport patients or specimens in a hospital by assigning costs to the messenger and to the time during which a request is in queue. In this problem, non-routine requests are superimposed on top of routine, scheduled requests. The authors also calculate the number of servers required so that a given percentage of requests do not exceed a given wait time and the average number of patients in the queue do not exceed a given threshold.

Assuming a phase-type service distribution, Gorunescu et al. (2002a) assign costs based on a base stock inventory policy. In this pure loss model, there is a holding cost associated with an empty bed, a penalty cost associated with each patient turned away, and a profit assigned to each day a bed is occupied.

Khan and Callahan (1993) incorporate advertising into their model to control the demand for laboratory services. For each staffing level, they determine the number of clients that would maximize profits. They then choose the staffing level with maximum profits and apply the necessary amount of advertising that would attract the desired number of clients. The model assumes that clients would leave without service if they wait above a certain amount of time.

Rosenquist (1987) chooses staffing capacity in an outpatient radiology service with a limited waiting area by minimizing cost. He suggests scheduling patients when possible and segregating patients based on expected examination duration. Such measures would reduce variability and decrease expected waiting times.

Gorunescu et al. (2002b) use backup beds (only staffed during peak demand) to reduce the probability of patient turn-away at a marginal cost. The model assumes a phase-type service distribution.

2.2.7 Appointment Systems

Compared to systems without appointments, systems with appointments reduce the arrival variability and waiting times at the facility. However, it is important to note that systems with appointments require patients to wait outside the facility. Of course, because it is not at the facility, this waiting can be productive time and therefore has lower cost to the patient. (Plus, they do not occupy space in the facility's waiting rooms.) A key issue has been to reduce patient waiting times without causing a significant increase in doctor idle time, a significant cost for the healthcare facility.

Bailey (1952, 1954) proposes (a) appointment interval and (b) consultant arrival time as two variables that determine the efficiency of an appointment system. In order to find a balance between patient wait time and consultant idle time, first determine the relative values of patient time and consultant time. The ratio of the total time wasted by all patients to the consultant's idle time should equal the value of the consultant's time relative to the patients'. He chooses to assign individual appointment times at intervals equal to the average patient processing time and finds that the consultant should arrive at the same time as the second patient.

Brahimi and Worthington (1991) design an appointment system to reduce the number of patients in the queue at any time, and reduce patient waiting time without significantly increasing doctor idle time. They explore the effect of patients who do not show up for their appointments. The clinic starts out with a certain number of patients waiting and a maximum number of patients allowed at any time.

Vasanawala and Desser (2005) reported that a radiology department has some time slots scheduled for routine radiology analysis. Emergency requests may require rescheduling of scheduled requests. Given a 1% or 5% probability of rescheduling, the authors use queuing theory to determine how many scheduled slots to leave empty during routine scheduling.

DeLaurentis et al. (2006) point out that patient no-shows without cancelling appointments could lead to waste of resources. They propose implementing short-notice appointment systems based on a queuing network analysis tailored to the realities of any particular outpatient clinic. Their approach assumes the availability of a certain number of staff who can be distributed amongst the different stations of the queuing network in several combinations. A combination is chosen based on its resulting utilization per station and expected patient length of stay in clinic. The implementations of these ideas did not improve the appointment system, a failure which they attribute to the clinic using many visiting doctors and the patients being unable to schedule visits with their primary care physician at short notice.

2.2.8 Bottlenecks

In a queuing network, there are several nodes at which services are dispensed. A patient may have to go through several nodes and thus several queues in order to obtain the desired service. In the context of appointment systems, we can expect nodes where the ratio of demand to available service capacity is relatively high to become bottlenecks. Such bottlenecks would have high utilization and increase overall patient waiting times even though other nodes may have low utilization.

Albin et al. (1990) find the bottlenecks at the Hurtado Health Centre appointment clinic by collecting data and analyzing it using QNA, the queuing network analysis software program. Though their model deviates appreciably from assumptions, they are able to find the bottle necks by identifying the nodes where wait times are longest. They then reduce overall waiting time by offering common-sense recommendations on a node-by-node basis.

2.2.9 System Size

Hall et al. (2006) said we can distinguish between three different scales. The smallest scale is the department, "a unit within a larger Centre oriented toward performing a single function, or a group of closely related functions." The next larger scale organization is the health care Centre, which is a group of proximate, coordinated departments amongst which patients can flow. The largest scale that we consider is the regional health system, a hierarchy of facilities with the most routine services provided by local clinics and the most specialized, resource-intensive services provided at a few regional facilities.

Hall et al. (2006) also present a macro system scale that considers the life cycle of an individual's state of wellness and his interactions with the healthcare system throughout a lifetime.

Kao and Tung (1981) investigate the redistribution of hospital beds amongst the inpatient departments of a hospital. First, a baseline patient capacity is chosen for each department. Additional beds are then allocated to departments in a manner to minimize patient overflows from one department to another. Forecasts are used to determine both the baseline bed allocation and the anticipated patient demand in order to minimize overflow.

Blair and Lawrence (1981) investigate a regional hierarchy of burn care facilities where excess demand at one facility is absorbed by other facilities in the same region and overflows at one region are absorbed by other regions. Worthington (1991) considers the coordination of patient flow from outpatient clinics to inpatient clinics.

Koizumi et al. (2005) model the mental healthcare system as a chain of facilities including acute hospitals, extended acute hospitals, and residential facilities and supported housing.

CHAPTER THREE

3.0 RESEARCH METHODOLOGY

3.1 INTRODUCTION

Data on arrival times, time service begins, time service ends, and departure time of out-patients was collected over a week (5 working days from Monday to Friday). This data will enable me to obtain the arrival rate, the service rate, and the traffic intensity of the patients using results from the birth and death model (which is synonymous to arrival and departure).

3.2 Population/Sample

This research evaluates the causes of excessive queuing in the outpatients' department of the Seventh Day Adventist Hospital which results in patient backlogs and long waiting times. The outpatient department of Seventh Day Adventist Hospital in this study operates from 8:00 am to 2:00 pm during weekdays (Monday to Friday) with two different types of patient visits: appointment patients and new patients. Appointment patients have priority over the new patients since most of them are actually presenting their laboratory results obtained the previous day to the doctor. They therefore generally spend less time at the hospital.

New patients who end their time at the outpatients department are those considered for this study. Also, outpatients who require laboratory services are not also considered for this study. This group may have to come back the next day to complete the system due to the mounting numbers of patients queuing to see the doctor.

3.3 Data Collection Procedure

Data was collected solely from the outpatient department of Seventh Day Adventist Hospital (SDAH). For any patient who enters the system (hospital) for treatment, he/she would have to go through series of stages. The patient joins queues at each of the stages to be attended to.

Like other health centres, a patient will have to go through the Records section (arrival point), health insurance section, History section and consulting rooms. When a patient is not covered by health insurance, he/she joins three (3) queues. After seeing the doctor, outpatients proceed to either the dispensary and/or the laboratory. Patients who Visits the laboratory are not considered in this research. Data was collected from all the sections of the outpatient department Monday to Friday.

Personnel were positioned at each of these stages to collect data on the following:

i. The number of arrivals entering the system
ii. Arrival rate of patients per hour
iii. The service time of each patient observed at each of the sections

Secondary data on outpatients during the one week period was also collected from the records section of the OPD to validate the primary data collected by the personnel.

3.4 Formulation of Problem

Here some formulae are derived and others quoted to use in the analysis of the data to help determine the set objectives.

3.4.1 Model Specification

The m/m/1 Queue (Single-Channel Queuing System). In this queuing system, the customers (patients) arrive according to a Poisson process with rate (λ). The time it takes to serve every customer is an exponential random variable with parameter (μ).

The service times are mutually independent and further independent of the inter arrival times. When a customer enters an empty system, his service starts at once and if the system is non-empty, the incoming customer joins the queue.

When a service completion occurs, a customer from the queue if any enters the service facility at once to get served.

The process $(X(t), t \geq 0)$ is a birth and death process with birth rate

$\lambda_i = \lambda \forall i \geq 0$ and death rate $\mu_i = \mu \forall i \geq 1$

The single-channel queuing system (m/m/1) with Poisson arrivals and exponential service rate and arrivals are handled on a first come first serve (FCFS) basis. In this queuing system, the average arrival rate is less than the average service rate (i.e. $\lambda < \mu$).

3.5 Performance Measures

Performance measures are used to gain useful information about waiting line systems. These measures include:

1. The average number of patients waiting in line and in the system. The number of patients waiting in line can be interpreted in several ways. Short waiting lines can result from relatively constant patient arrivals (no major surges in demand) or from

the organization's having excess capacity (many servers). On the other hand, long waiting lines can result from poor server efficiency, inadequate system capacity, and/or significant surges in demand.

2. The average time patients spend waiting, and the average time a patient spends in the system. Patients often link long waits to poor-quality service. If too much time is spent in the system, patients might perceive the competency of the service provider as poor.

3. The system utilization rate. Measuring capacity utilization shows the percentage of time the servers are busy. Management's goal is to have enough servers to assure that waiting is within allowable limits but not so many servers as to be cost inefficient.

These measures are calculated for two different waiting line models: the single-server model and the multi- server model. The easiest waiting line model involves a single-server, single-line, single-phase system.

3.5.1 Assumptions

The following assumptions are made when we model this environment:

1. The customers are patient (no balking, reneging, or jockeying) and come from a population that can be considered infinite.
2. Patient arrivals are described by a Poisson distribution with a mean arrival rate of λ (lambda). This means that the time between successive patient arrivals follows an exponential distribution with an average of $\frac{1}{\lambda}$

3. The patient service rate is described by a Poisson distribution with a mean service rate of μ (mu). This means that the service time for one patient follows an exponential distribution with an average of $\frac{1}{\mu}$
4. The waiting line priority rule used is first-come, first-served.
5. The system is in a steady state. The mean arrival rate is the same as the mean departure rate.
6. The mean arrival rates and service rates are constant.

Using these assumptions, we can calculate the operating characteristics of a waiting line system using the following formulae.

3.5.2 Mean Arrival Rate (λ)

Let n = the number of patients who entered the system between 8:00am to 2pm each day.

h = the number of hours between (8:00am to 2pm) each day

Hence the mean arrival rate (λ) is given by:

$$\lambda = \frac{n}{h} \qquad (1)$$

The mean arrival rate (λ) will have units of arrivals/hours

Then, the mean time between arrivals (mean inter arrival time) is given as $\frac{1}{\lambda}$

3.5.3 Mean Service Rate (μ)

The mean service rate (μ) is determined for patients at each of the four stages at the OPD by placing data collection personnel at each of the points to collect the data on service rate per hour

3.6.0 Single Server Models

The following formulas are developed for the single-channel queuing system (m/m/1) with Poisson arrivals handled on a first come first serve (FCFS) basis and exponential service rate.

The traffic Intensity (ρ) called the quantity of work brought to the system per unit time (the probability of queuing on arrival) is given as:

$$\rho = \frac{\lambda}{\mu} \qquad (2)$$

Also, the probability of no patient in the system (Probability no queuing on arrival) P_o is given as:

$$P_0 = 1 - \rho \qquad (3)$$

Then the probability of having to wait at the hospital is given as:

$$P(w) = 1 - \rho_0 \qquad (4)$$

The average number of patients on the queue at any given time t is given as:

$$\frac{\rho^2}{1-\rho} \qquad (5)$$

The average number of patients waiting to be served at any time t is given as:

$$\frac{1}{1-\rho} \qquad (6)$$

The average number of patients in the system is given as

$$\frac{\rho}{1-\rho} \qquad (7)$$

The average time in queue (before service is rendered) is given as

$$\frac{\rho}{\mu(1-\rho)} \qquad (8)$$

The average time in the system (on queue and receiving service) is given as

$$\frac{\rho}{\mu(1-\rho)} \qquad (9)$$

The probability that there are n patient in the system is given as:

$$(1-\rho)\rho^n \qquad (10)$$

3.7 Single Channel Multi- Server Models

At the consultation section, a single channel multi server-model is used for the determination of the performance measures. This is because there is more than one doctor carrying out consultations at the unit.

Hence the following under stated queuing formulae are used.

Let S= number of servers

L = mean number in the system (including the one being served)

L_q = expected number in the waiting line (expected length of the waiting line).

W_q = mean waiting time

Number of patients waiting in queue is given as;

$$L_q = \frac{\left(\frac{\lambda}{\mu}\right)^2 \frac{\lambda}{s\mu}}{s!\left(\frac{\lambda}{s\mu}\right)^2} \rho_0 \quad (11)$$

Average Patients waiting time is calculated as;

$$W_q = \frac{L_q}{\lambda} \quad (12)$$

Average number of Patients in the system:

$$L = L_q + \frac{\lambda}{\mu} \quad (13)$$

Average Patient Time spent in the system

$$W = W_q + \frac{1}{\mu} = \frac{L}{\lambda} \quad (14)$$

Server Utilization Factor:

$$\rho = \frac{\lambda}{s\mu} \quad (15)$$

3.8 Limitations of Queuing Models

Queuing models have several limitations. Most of these limitations are the basic assumptions for application of queuing models. Some of the limitations of queuing models are enumerated below:

- Takes average of all variables rather than the real numbers itself
- Assumes steady state
- Based on assumption that service time is known
- Service times are independent from one another
- Service rate is known
- Service rate is greater than arrival rate
- Arrivals are served on first come first serve basis
- Service times are described by the negative exponential probability distribution

CHAPTER FOUR

4.0 RESULTS AND DISCUSSION

4.1 Introduction

This chapter looks at the Presentation of data collected for this work, analysis and presentation of results of findings. This is presented in a form of tables, graphs mathematical calculation using the queuing models in chapter three.

4.2 Presentation and preliminary analysis of data

The SDA hospital on daily bases has servers who attend to patients at each of the OPD sections. These sections include the Records (arrival point)/health insurance section, History section, consulting rooms and the dispensary. Normal outpatient service at this hospital starts from 8:00am to 2:00pm daily. This translates into 6 working hours per day for outpatient attendance. The average daily attendance of patients at this hospital is computed and use for the analysis of the queuing system at the SDA hospital. The data collected during the one week period as well as patients daily average arrival rates is presented in the following table.

Table 4.1 Daily Patients attendance at the General OPD sections

DAY	PATIENTS DAILY ATTENDANCE		
	New patients	Old patients	Total (N)
Monday	91	124	215
Tuesday	75	98	173
Wednesday	72	108	180
Thursday	52	105	157
Friday	70	103	173
Total	**360**	**538**	**898**
Average Daily Records	**72**	**108**	**180**

4.2.1 Computation and analysis of the performance measures

The average number of patients who attends the SDA hospital daily is 180 made up of 72 patients who are attending the hospital for the first time (required new cards) and 108 old patients who already have record/files at the hospital. This represents 12% new and 88% old.

The daily arrival rates (λ) at each of the four sections of the OPD are computed in the following table below.

Table 4.2 Daily Arrival Rates at the General OPD sections.

DAY	ARRIVAL RATE (PATIENT/HR).			
	Records (Arrival Point) /Health Insurance Section	History Section	Consulting Rooms	Dispensary
Monday	32	33	35	39
Tuesday	31	32	34	36
Wednesday	28	30	31	34
Thursday	30	32	34	35
Friday	27	29	30	35
Total Service rate	**148**	**156**	**164**	**179**
Mean Service rate (λ)	**30**	**31**	**33**	**36**

The daily service rates at each of the four sections of the outpatient department at the Tamale SDA hospital during the one week period of the data collection are recorded and computed in table 4.3 below.

Table 4.3 Daily Service Rate at the General OPD sections.

DAY	SERVICE RATE (PATIENT/HR).			
	Records (Arrival Point) /Health Insurance Section	**History Section**	**Consulting Rooms**	**Dispensary**
Monday	33	35	36	41
Tuesday	32	34	36	40
Wednesday	30	32	34	37
Thursday	32	34	35	36
Friday	29	30	39	41
Total Service rate	**156**	**165**	**180**	**195**
Mean Service rate (μ)	**31**	**33**	**36**	**39**

The mean arrival rates (λ) and the mean service rate (μ) at the various sections of the OPD are summaries in the diagram below:

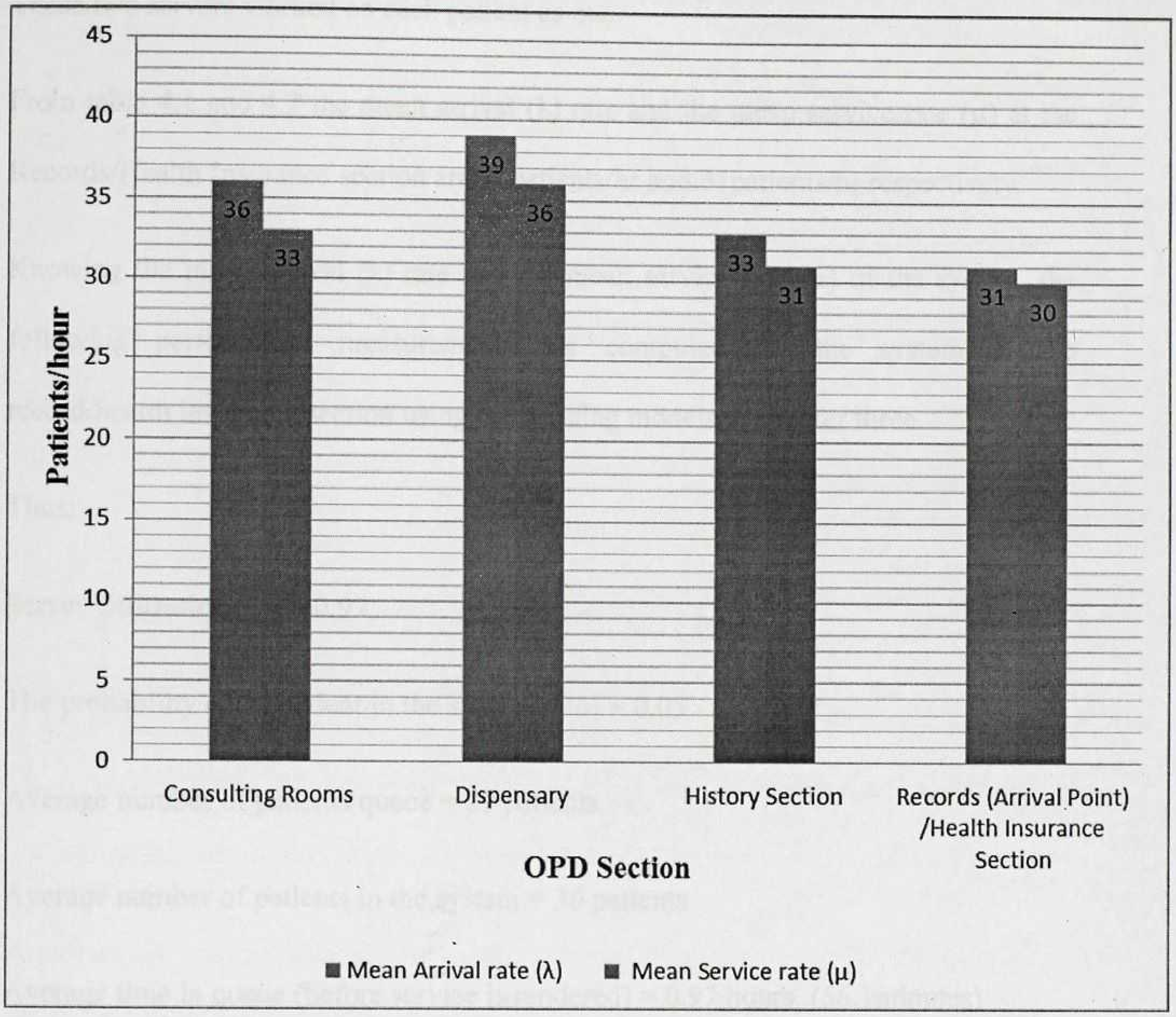

Figure 4.1: Mean Arrival Rates and Mean Service Rates

Clearly, the diagram shows that the Mean service rate (μ) at each of the sections is greater than the mean arrival rate (λ) at each section. This implies that the queue will come to an end at a particular time.

4.2.2 Performance Measures at the Records/Health Insurance section

The number of servers at the records/ health insurance section of the OPD are two (2). These two servers worked on each patient as one.

From table 4.1 and 4.2 the mean arrival (λ) rate and the mean service rate (μ) at the Records/Health Insurance section are 30patients/hr and 31patients/hr respectively.

Knowing the mean arrival (λ) rate and the mean service rate (μ) of the system, the following performance measurements are computed for the system at the record/health insurance section using the queuing models in chapter three.

Thus:

Server Utilization (□) = 0.97

The probability of no patient in the system (□o) = 0.03

Average number of patients queue = 29 patients

Average number of patients in the system = 30 patients

Average time in queue (before service is rendered) = 0.97 hours. (56.1minutes)

Average time of the system (On queue and being served) = 1.00 hour. (60 minutes)

Average time of service = 0.03 hours (1.8 minutes)

The result shows that the mean server capacity utilization at this section is 97%. This means that the servers at the records unit will be busy 97% of their time. Conversely, the servers at the records will be idle for 3% of the working hours.

The average number of patients waiting in the queue for service at the OPD at any given time is found to be twenty nine (29) patients and the average number of patients in the system (patients in queue plus patients being served) is thirty (30).

The probability of a patient waiting at this unit before service is rendered is 0.97 and the average waiting time of a patient in queue (before service is rendered) is 0.97 hour, (ie 58.1minutes).

Finally, the average waiting time in the system (on queue and being served) is 1.0 hour, (ie 60 minutes) and the average time of service is 0.03hr (i.e 1.9 minutes).

4.2.3 Performance Measures at the history section

There is only one (1) server at the history section of the OPD.

From table 4.1 and 4.2 the mean arrival rate (λ) and the mean service rate (μ) of patients at the history section are 31patients/hr and 33patients/hr respectively.

Knowing the mean arrival (λ) rate and the mean service rate (μ) of the system, the following performance measurements are computed for the system at the History section using the queuing models in chapter three.

Thus;

Server Utilization (□) = 0.94

The probability of no patient in the system (□o) = 0.06

Average number of patients queue = 15 patients

Average number of patients in the system = 16 patients

Average time in queue (before service is rendered) = 0.47hours. (28.1minutes)

Average time of the system (On queue and being served) = 0.50hours. (30minutes)

Average time of service = 0.03hours (1.8minutes)

The result shows that the capacity utilization of the servers is 94%. This implies that the servers at the records will be idle for 0.6% of the working time on Monday.

The average number of patients waiting in the queue for service at the history section at any given time is found to be fourteen (15) patients and the average number of patients in the system (patients in queue plus patients being served) is fifteen (16). The probability of a patient waiting at this unit before service is rendered is 0.94 and the average waiting time of a patient before service is 0.47 hours, (ie 28.1minutes).

Finally, the average waiting time in the system (on queue and being served) is 0.50 hour, (ie 30 minutes) and the average time of service is 0.03hr (i.e 1.8 minutes).

4.2.4 Performance Measures at the Consulting Room

There are three doctors doing consultation in three consulting rooms on patients at this unit of the OPD each day. The research assumed that these three doctors report and start work at the same time and their service rates of are the same (ie $\mu_1 = \mu_2 = \mu_3$). Where μ_1, μ_2 and μ_3 are the service rates of doctor 1 in consulting room 1, doctor 2 in consulting room 2 and doctor 3 in consulting room 3,

The mean arrival rate (λ) at the consulting section of the OPD is 33patients/hour and the total service rate (μ) of the three doctors at this section is found to be 36patients/hr (ie μ=32patients/hr). This implies that an average of 12patients/hr is attended to by each of the three doctors. The following performance measurements are computed for the system at consulting section.

Thus;

Server Utilization (□) = 0.92

The probability of no patient in the system (□o) = 0.08

Average number of patients queue = 10 patients

Average number of patients in the system = 11 patients

Average time in queue (before service is rendered) = 0.92hours. (ie 55.2minutes)

Average time of the system (On queue and being served) = 1.0hour (ie 60minutes)

Average time of service = 0.08hours (ie 5minutes)

The result shows that the capacity utilization of the servers (server busy time) is 92% and will be idle for 8% of the working time. The average number of patients waiting in the queue for service at the consulting room at any given time is ten (10) patients and the average number of patients in the system (patients in queue plus patients being served) is eleven (11).

The average waiting time of a patient before service at the consulting section is 0.92 hours, (ie 55 minutes) and the probability of a patient waiting before consultation service is rendered is 0.92.

4.2.5 Performance Measures at the Dispensary section

The number of servers at the dispensary section of the OPD each day is 2. These servers worked as if it is one server attending to a patient at a time. The mean arrival rate (λ) at the Dispensary section is 36patients/hour and the mean service rate (μ) at this section is found to be 39patients/hour

Using the queuing models in in chapter three, the following performance measures are computed for the system at the Dispensary section.

Thus;

Server Utilization (□) = 0.92

The probability of no patient in the system (ρ_o) = 0.08

Average number of patients queue = 11 patients

Average number of patients in the system = 12 patients

Average time in queue (before service is rendered) = 0.310hour (18.5minutes)

Average time of the system (On queue and being served) = 0.33 (20minutes)

Average time of service = 0.03hours (1.8 minutes)

The result shows that there is 92% utilization of the servers at the dispensary section. The servers will be idle for 8% of the working time. The average number of patients waiting in the queue for service at this section at any given time is eleven (11) patients and the average number of patients in the system (patients in queue plus patients being served) is twelve (12).

The average waiting time of a patient before service at the dispensary section is 0.31 hours, (ie 18.5 minutes). The probability of a patient queuing before service is rendered at the dispensary is 0.92.

The computed performance measure of the system (Tamale SDA Hospital OPD) is summarised in the following table.

Table 4.4: Summary of Performance Measure at the OPD.

OPD Sections	**λ**	**μ**	**ρ**	**Average Number in Queue (L_q)**	**Average Number in the system (L)**	**Average waiting time of a patient in queue (W_q)/hr.**	**Average time spent in the system/ hr.**	**Average time of service (and being served)/ Min**
Records/ Health Insurance Section	31	30	88	29	30	0.97	1.00	1.80
History Section	33	31	94	15	16	0.47	0.50	1.80
Consulting Rooms	36	33	91	10	11	0.92	1.00	0.08
Dispensary	39	36	88	11	12	0.31	0.33	1.80
Totals				**65**	**69**	**2.67**	**2.83**	**5.48**

CHAPTER FIVE

5.0 CONCLUSIONS AND RECOMMENDATION

5.1 Introduction

This chapter looks at the summary of the entire study, interpretation of results and conclusions of the research. It also looks at the limitations of the research, recommendations and directions for future research work.

5.2 Summary and Interpretation of Results:

The results shows that the server utilization at the Records (arrival point)/health insurance section, History section, consulting rooms and the dispensary is 88%, 94%, 91%and 88% respectively. This implies that the probability of a patient queuing at each of the section at the OPD is 0.88, 0.94, 0.91 and 0.88 respectively.

Though the service in the SDA hospital is not 100% efficient, the resource utilization is very high thereby keeping the servers busy during the working hours.

The average number of patients waiting in queues for service at the OPD at any given time of the working session is 65. The average number of patients in the system (On queue and being served) at any given hour during work time is found to be 69. It then follows that there will always be queues since the average time in the queue system (both on queue and receiving service) is greater than the average time in queue before service is rendered

The average amount of time spent waiting in queue at each of the four sections at the OPD ie Records (arrival point)/health insurance section, History section, consulting rooms and the dispensary before service is rendered are 0.97, 0.47, 0.92 and 0.31 hours respectively. This brings the average waiting time at OPD to 2.67 hours per person. This implies that a patient will have to wait for 2.67 hours before receiving total service at the OPD.

Also the total waiting time of patients in the system (On queue and being served) is found to be 2.83hours (ie 169.8minute).

The research shows that the queues and waiting times are too large and would probably result in decreased satisfaction of the patient and increased employee workload.

5.3 Limitations of the Study

Despite its contributions, this study also has its own limitations. Most particularly, I recognize that the data was gathered within a one week period from Monday to Friday in a single hospital. This may limit the generalization of the findings. This is mainly because of limited resources and time to undertake the study on a wider scale.

Moreover, only the outpatient department of the hospital was considered in this study. That is, a queuing network model should be considered to analyze the patient flow of the entire hospital.

Additionally, this study takes into consideration the actual waiting time but ignores the effect of perception of waiting time on patient satisfaction.

5.4 CONCLUSION AND RECOMMENDATIONS

In a period of major change in the health care environment, enhancing satisfaction is becoming increasingly important because satisfaction is recognized as a measure of quality. Knowledge of the use of queuing model to determine system parameters is of value to health providers who seek to attract, keep and provide quality health care to patient in the ever-competitive “marketplace”.

Providing patients with timely access to appropriate medical care is an important element of high quality care which invariably increases patient satisfaction, when care is provided is often as important as what care is provided

Queuing theory is a useful statistical technique for solving peculiar problems. Its applications in an organization are indispensable. Its use has been validated in industrial setting, retail sector and in service settings such as telecommunications but its adoption and use in healthcare setting is lagging behind other sectors. In health sector it is mainly used in Emergency Department (ED) wait line and staffing studies, analysis of queues in outpatient and ambulatory care settings and for disaster management.

It can be used in inpatient, outpatient, physician office, public health, facility and resource planning, emergency preparedness, mental health, long term care, dispensary, inventory control as well as public health.

Knowledge of queuing theory can help service managers to make decisions that increase the satisfaction of all concerned parties (patients/customers, employees, and management).

The queuing problems encountered at the Tamale Seventh Day Adventist Hospital (SDAH) are similar to what is encountered in older centres as well as other government hospitals across the country. Excessive waste of time in the hospitals or health centres may lead to patients' health complications and in some cases eventual death which may be avoided.

Three main causes have been identified as the contributing factors towards excessive patient's waiting time, namely: the registration time, insufficient number of counter service staff and insufficient number of doctors.

To ensure satisfaction of all parties at the hospital, the following recommendations are made for the management of the Tamale SDA hospital.

1) Hospital authorities should make efforts to educate public on the start hours of OPD section. This is because majority of the patients report at the OPD between the hours of 6:00am to 8:00am which is outside the start time of 8:00am. This creates the long waiting times before service time began.
2) It is recommended that more doctors and nurses be deployed to these sections so as to convert the single-channel queuing units to multi-channel queuing units. This will offer a speedy service to patients on arrival particularly during the peak hours (ie from 8:00am to 11:00am.
3) Again, paramedical officers should be deployed to the hospital. This will take care of patients' preliminary tests or service before they see the doctors. This will reduce the service time spent by the doctors in attending to patients and hence the service efficiency.
4) There should be a constant system monitoring at the hospital (ie actual arrival and start time of the health personnel at work, tracking number of patients waiting by location). This will ensure that doctors and other health personnel's report and start work at the right time.
5) Health education programs should be introduced at the OPD to engage patients waiting for service. Patients will see this as part of the hospital procedures and therefore will not feel bored.

6) Finally, a further study is recommended to cover some selected hospitals across the region or the nation and for a longer period of time to enable generalization of the findings of the study.

REFERENCES:

1) Adedayo, O. A., Ojo, O., Obamiro, J. K., (2006) *Operations Research in Decision Analysis and Production Management*. Lagos: Pumark Nig. Ltd.
2) Adele, M., Barry, S.,(2005) *Modeling patient flow in hospital using queuing theory*, Unpublished Manuscript.
3) Adeleke, R.A., C.E. Adebiyi, and O. Akinyemi, (2006) "Application of Queuing theory to Omega Bank PLC, Ado- Ekiti, Nigeria".
4) Adeleke, R.A., O.D. Ogunwale, and O.Y. Halid, (2009) "Application of Queuing Theory to Waiting Time of Out-Patients in Hospitals". *Pacific Journal of Science and Technology.*
5) Au-Yeung, S.W.M., Harrison, P.G., & Knottenbelt, W.J., (2006), A queuing network model of patient flow in an accident and emergency department, Department of Computing, Imperial College of London.
6) Babes, M., Serma, G. V., (1991) Out-patient Queues at the Ibn-Rochd Health Centre, *Journal of Operational Research.*
7) Belson, G. V., (1988) Waiting Times Scheduled Patient in the Presence of Emergency Request, *Journal of Operational Management.*
8) Bretthauer, K.M. & Cote, M.J., (1998), "A model for planning resource requirements in health care organizations",

9) Buhaug, H., (2006), "Long Waiting List in Hospitals", British Medical Journal.

10) Bunday, B.D., *An Introduction to Queuing Theory*, New York: Halsted Press, 1996

11) Carter, M.W., (2006), Operations research in healthcare or who let the engineers into the hospital? RCHE Pioneers in Healthcare Engineering Colloquium.

12) Chaska, L. N. et al., (1980) Use of Medical Services and Satisfaction with Ambulatory Care among a Rural Minnesota Population, *Public Health Reports*.

13) Davis, M.M., Heineke, J., (1998) *How Disconfirmation, Perception and Actual Waiting Times Impact*, Unpublished Manuscript.

14) Davis, M.M., Vollmann, T.E., (1990) A framework for Relating Waiting Time and Customer Satisfaction in a Service Operation, *Journal of Services Marketing*.

15) Churchman, G.W., R.C. Ackoff, and F.C. Arnoff, (1957) *Introduction to Operation Research*. John Willey and Sons: New York, NY.

16) Cote, J.M., (2000), Understanding Patient Flow, Decision Line,

17) Gorney, L, (1981) *Queuing Theory: A Problem Solving Approach*, New York: Petrocelli Booko, Inc.

18) Green, L.V., Soares, J., Giglio, J.F. & Green, R.A., (2006), "Using queuing theory to increase the effectiveness of emergency department provider staffing", Academic Emergency Medicine Journal, Volume 13.

19) Green, L.V. & Nguyen, V., (2002), "Strategies for cutting hospital beds: The impact on patient service", Health Services Research.

20) Griffiths, J.D., Lloyd, N.P., Smithies, M. & Williams, J., (2006), "A queuing models of activities in an intensive care unit", IMA Journal of Management Mathematics.

21) Gross, D., Harris, C.M., (1985) *Fundamentals of Queuing Theory*. New York: John Wiley, Second Edition.

22) Hall, W. R., (2006) Patient Flow: The New Queuing Theory for Healthcare, OR/Ms, *Today*.

23) Ivalis, S., Millard, P., (2003) Health care modeling operating the black box, *British Journal of Health Care Management*.

24) Jr., R.N.A., & Wilson, J.P., (2001), Queuing theory and customer satisfaction: A review of terminology, trends and applications to dispensary practice, Hospital Dispensary.

25) Kandemir-Caues, C., Cauas, L., (2007) An Application of Queuing Theory to the Relationship between Insulin Level and Number of Insulin Receptors, *Turkish Journal of Biochemistry*.

26) Katz, K., Larson, K., Larson, R., (1991). Prescription for the waiting in line Blues: Entertain, Enlighten and Engage", *Sloan Management Review* (winter)

27) Larson, R.C. & Odoni, A.R., (1999), "Chapter 4: Introduction to queuing theory and its applications", Urban Operations Research, Massachusetts Institute of Technology.

28) Mackay, M. & Lee, M., (2005), "Choice of models for analysis and forecasting of hospital beds", Healthcare Management Science.

29) Medhi, J., (2003) *Stochastic Models in Queuing Theory.* Amsterdam: Academic press, Second Edition.

30) Nosek, A.R., Wislon, P. J., (2001) Queuing Theory and Customer Satisfaction: A Review of Terminology, Trends and Applications to Pharmacy Practice, *Hospital Pharmacy.*

31) Obamiro, J.K., (2003) *Application of Queuing Model in Determining the Optimum number of Service Facility in Nigerian Hospitals*, M. Sc. Project submitted to Department of Business Administration, University of Ilorin.

32) Ozcan, Y.A., (2006), Quantitative methods in health care management; Techniques and applications, Firs edition, Jossey -Bass Publications.

33) Preater, J., (2001), A bibliography of queues in health and medicine, Mathematics Department, Keele University.

34) Preater, J., (2002), "Queues in health", Healthcare Management Science.

35) *Queuing Theory and Patient Satisfaction: An Overview of Terminology and Application* ... 11

36) Singh, V.,(2007) *Use of Queuing Models in Health Care,* Department of Health Policy and Management, University of Arkanses for medical science,.

37) Sitzia J., Wood N., (1997) Patient Satisfaction: A Review of Issues and Concepts, *Soc Sci Med,*

38) Taha, H.A. 1987. *Operation Research: An Introduction, Fourth Edition.* Macmillan Publishing: New York, NY. pp 595-671.

39) Taylor, S., Waiting for Service, (1994): The Relationship between Delays and Evaluation of Service, *Journal of Marketing,*

40) Weber, D.O., (2006), "Queue Fever", Hospitals & Health Network.

41) Young, G.J., Meterko, M., Desal, R. K., (2000) Patient Satisfaction with Hospital Care: Effects of Demographic and Institutional Characteristics.