Queuing Network Analysis and Optimal Bed Determination: An Evaluation of Nigerian University Teaching Hospital Emergency Department

Obamiro John Kolade (PhD)
Department of Business Administration,
Lagos State University, Ojo
johnobamiro@gmail.com

Abstract

This study uses a decomposed queuing network model to determine optimal beds to serve an emergency department where patients have low tolerance for delay. M/M/∞ and M/M/C of queuing models were used to analyse the data obtained from LUTH Emergency Department (ED) for a period of 18 months. The findings reveal that the ED requires at least 64 beds to provide efficient emergency services to the existing population. It was suggested that improving patient flow and reducing waiting time for patient can be achieved through the following ways: (1) managing the arrival rate by different appointment methods such as online appointment management system to spread arrivals and call in appointment system; (2) increasing the number of servers; (3) optimizing the service rate.

Keywords: Queuing Network Model; Hospital; Emergency Department; Overcrowding; Waiting Time; Optimal Beds.

Introduction

Nigerian public hospitals and all over the world are faced with the challenge of allocating limited resources, such as beds, personnel and other medical resources available for randomly arriving patients. One of the hospital units that have been seriously affected is the Emergency Department (ED). What account for this recent experience is the increasing rate of population and its demand for emergent medical emergent services. Overcrowding in ED which prevents prompt delivery of medical services may have negative effect on the quality of health care. The inability of the ED resources to hand rapid changes in demand for care services results to high levels of congestion in the ED. Scarcity is observed both in resources unable to change in the short term, such as nurses and physicians, as well as resources that are rather fixed, such as bed capacity. Congestion or waiting is experienced due to random arrivals and length of stay of patients in the ED. It has been an issue all over the world to manage patient waiting time, deliver prompt care and provide patient satisfaction in ED and Nigeria medical situation is not an exception. So, health care systems have been challenged in recent years to deliver services to all the patient and high quality services with limited resources without delay **(**Malik and Belwal, 2016). In an ideal hospital with perfect information about patients' arrivals and length of stay, it would be feasible to allocate the exact amount of resources required in ED (Hajnal and Zsuzsanna, 2015; Diakogiannis, 2014), using queuing model.

Researches show that in the emergency department(ED) and intensive care units (ICU) of the Public Teaching Hospitals (Lewin Group, 2004; IOM, 2006; Av-yeung, Herrison and Knottenberg, 2006; Au-Yeung, Harder and McCoy, 2009; Green, Soares, Giulio and Green, 2006; Delia 2007; Bruin, Rossum, Visser and Koole, 2007; Geranmayeh and Iyer, 2008; Olorunsola, Adeleke and Ogunlade, 2014; Aboukanda and Latif, 2014; Diakogiannis, 2014; Hajnal and Zsuzsanna, 2015), patients experience longer waiting times to be admitted to or diverted from a unit (as a bottleneck unit) as it reaches capacity thereby reducing healthcare access to the public, and increasing operational cost to hospitals because of the associated inefficiencies. The long waiting lists have become symbols of the inefficiency of hospitals services all over the world, particularly in publicly funded hospitals (Gauld, 2000). Overcrowded emergency department (ED), intensive care units, prolonged waiting times, patient care delays, refused or diverted admissions and scare resources are common themes in large public hospitals (Derlet and Richards, 2000; Henry, 2001; Shmueli, Springe and Kaplan, 2002; Lewin Group, 2004; IOM, 2006; Av-yeung, et al, 2006; Green, Soares, Giulio and Green, 2006; Delia 2007; Bruin, Rossum, Visser and Koole, 2007;

Geranmayeh and Iyer, 2008; Obamiro2013; Zhu, Gong, Tang 2013 Olorunsola, Adeleke and Ogunlade, 2014; Qiang Liu, Xie and Liu 2014). Management of waits, delays and unclogging bottlenecks requires the assessment and improvement of flow between and among various departments, and optimal allocation of available resources in the entire hospital system. Current healthcare literature and practice indicate that waiting lists and congested patient flows are indeed made up of one of the most important problems in care industries (Belson, 2006). In order to improve performance in an environment as complex as a hospital system, the dynamics at work need to be understood, of which queuing theory and simulation provide an ideal set of instruments for such understandings (Creemers and Lambrecht, 2007).

Queuing theory is a mathematical science-based approach to analyze waiting lines and resources allocation (Bahadori, Teymourzadeh, Hosseini, and Ravangard, 2017) . Despite the fact that queuing theory was first applied to analyse telephone congestion by A. K. Erlang in 1913, it has found an effective and efficient application in health care setting. The major aim of queuing theory is to derive an analytical or mathematical model of customer needing service and use that model to determine waiting times and allocation of resources such as beds, doctors, nurses, healthcare machines, etc. (Malik and Belwal, 2016). Queuing models have been used in appropriate systems to determine the optimal supply of fixed resources necessary to meet a variable demand. For instance, Kyoung, Seong, Young and Yong (2017) applied **q**ueuing theory to the analysis of changes in outpatients' waiting Times in Hospitals introducing EMR. Gorunescu, Mcleen, Millard, (2002) used queuing model for designing, planning and staffing service units. Queuing theory was also used to determine capacity requirements especially beds (Green, 2002; McManus et al, 2004). Obamiro (2012) modeled patient flow and resource allocation in intensive care unit using queuing theory. He determined ICU efficiency and number of beds required to serve a given population. de Bruin (2007) determined the optimal bed allocation over the care chain given a maximum number of refused admissions. Agrawal (2010) used optimization model to determine the optimum number of doctors/service providers required during a service session. Olorunsola, et al (2014) applied queue theory to accurately model the flow of in-patient in hospital; they determined the optimal bed count and its performance measure.

Hospital queuing systems have several servers linked together. The output of one service centre (Record unit) may proceed to another service facility (Nurse unit or Consultation unit) for further processing, or return to facility already visited for further examination (Obamiro, 2011). A queuing model which composes of a set of linked queues called stations (i.e. multiple stations) is called a queuing network model. A queuing network is a connection of several queuing systems. A Queuing network is a version of queuing model that deals with analysis of patients or customers that require more than one service from different service facilities one after the other, and they have to queue up for service waiting for each of the servers. Patient flow in hospitals can be naturally modeled as a queueing network, where patients are the customers, and medical staffing, beds and equipment are the servers (Armony, 2015). Modeling complex systems using queuing network model allows us to better understand their behaviours, to estimate and ultimately to improve their performances (Osorio and Bierlaire, 2006). Considering a network of hospitals with different units, each unit is modeled as a specific queue and where it is the patient flow and optimal bed determination is the main focus. Emergency department system can be regarded as a network of queues and different types of servers where patients arrive, wait for a service, get a result and then go home or they are admitted to a hospital unit(Vass and Szabo, 2015). The analysis of patient and waiting times in this department will be enhance management decisions on resources (beds) allocation.

Jackson (1957, 1963), Koenigsberg (1958), made early and notable contributions and application of queuing network models. A Jackson model is probably the most researched and widely applied network model in various fields, including the healthcare field (Koizumi, 2002). The characteristics of a Jackson network are the same assumptions of a system of infinite capacity for all queues in series, except now, the customers visit the facilities in different order (and may not visit them all) (Hiller and Lieberman, 2005). A variety of queuing network frameworks has been developed to represent various system mechanisms. The system in a network model is characterized by: (a). An open or a closed system, and (b). Linkage of station (tandem, arbitrarily linked with or without feedback flow)(Koizumi, 2002). As in a single server station model, each station in network system owns characteristics including, (i) inter-arrival times, (ii) service times, (iii) the number of servers (iv) the maximum capacity of station, and, (v) queue discipline. In a queuing network model, waiting spaces between stations (i.e. part of each station's "capacity") are expressed as "buffer".

Institute of Medicine (2006) reports serious problems confronting hospital ED's across the nation as overcrowding (Delia, 2007). According to the Crowding Resources Task Force (2002), Emergency department crowding refers

to:“A situation in which the identified need for emergency services outstrips available resources in the ED. This situation occurs in hospital ED when there are more patients than staffed ED treatment beds and wait times exceed a reasonable period. Crowding typically involves patients being monitored in non treatment areas (e.g. hallways) and awaiting ED treatment beds or inpatient beds”.
Overcrowded emergency departments, prolonged waiting times, patient care delays and scarce resources are common themes in current teaching hospitals (Yoon, et al., 2003). Overcrowded ED has become a major barrier to receiving timely emergency care. Patients who present to EDs often face long waiting times to be treated and, for those who require admission, even longer wait for an inpatient hospital bed (Asplin, et al., 2003). The reasons of overcrowding are; “boarding” in the ED until an inpatient bed becomes available (Henry, 2001; Delia, 2007) and excessive wait times for medical care. These keep patients in ED beds, on stretchers placed in hallways, or in ‘observation’ areas with little if any regard for privacy, dignity or personal hygiene for hours and several days. These patients, while waiting to be transferred to the hospital, occupy the spaces for other patients who need “emergent evaluation of treatment” (Henry, 2001). The regular occurrence of ED overcrowding raises concern about the hospital sector’s ability to respond to a mass casualty event such as an oil disaster, plane crash, ethic war victims and fatal accidents involving vehicles (Delia, 2007). Overcrowded ED’s also create an environment where medical errors are more likely and overall quality of care is below its potential (ICAHO, 2004). For developed countries hospitals to prevent congestion, the ED often declares diversion, and asks ambulances to find alternate treatment facilities for their patient. As a result, “the ability of the hospital to provide emergency care to its community and serve its role in the emergency medical services (EMS) is lost” (Caglar, 2003). The overcrowding situation is more prevalent in Nigerian Teaching Hospitals because there is no provision for ambulance diversion. That is, in the event of natural disaster like ethnic war and oil vandals, patients are admitted on the floor or any available open space and some patients are rejected or send to primary health care centres of private hospitals that lack the required equipment and personnel.

In response to potential problems facing ED in Nigerian Teaching Hospitals, this research was undertaken to address the problem of bed capacity and model patient flow within the emergency department. To solve the problem of mismatches of larger supply and demand that result to overcrowded ED, we proposed a conceptual model of ED (similar to Asplin, et al., 2003 and Delia’s, 2007) that will be helpful to researchers, hospital administrators, policy makers to understand the cause, effect and proffer solution to ED overcrowding.

Capacity of an ED can be described in terms of staffing or beds. Capacity management in an emergent department involves optimal utilization of available resources (beds, doctors, nursing staff, equipment and operating theatre) (Volira et al. 2008). Quantitative analyses of optimal use of health care resources have often been conducted using queuing analysis (Green 2006). Researchers have also used single queuing models to analyze patient flow in ED. The models deal with the determination of optimal beds, nurses, doctors etc., are required to serve a given flow of patients in a clinical unit. Green et al (2007) applied queuing model for nurse staffing for time varying demand and (McManus et al. (2004) optimal number of beds in an intensive care unit (ICU) was determined using the model. In order to get exact or approximate estimation of performance of the ED is consists of several subunits and multiple patient classes, the queuing models of the proposed hospitals are developed using parametric decomposition approach of open queuing network. Parametric decomposition approach is the most popular approach to evaluate the performance of both infinite and finite capacity queuing networks and first researched by Jackson (Jackson (1957 and 1963). Although, it refined and applied in infinite capacity queuing models by authors such as Bitran and Tirupati(1988); Vandale, De Boeck and Callewier (2002); and Osorio and Bierlaire, (2007); Xie et al, (2007); Carolina (2009); Chow et al. (2011), Gupta (2007), Hall et al. (2006).

This study adopts the analytic approximation methods because it allows for a successful decomposition of complex systems of entire hospitals such as LUTH into sub-units (sub-multi-servers), where each unit has specific characteristics. Another reason is that analytic approximation methods reduce the dimensionality of the system under consideration. Decomposition methods achieve this by disaggregating the network into sub networks and analyzing each sub network in isolation (Osorio and Bierlaire, 2007). Existing decomposition methods have analysed simple sub networks consisting of single stations, pairs of stations and triplets. If not stated otherwise, the methods concern open finite capacity networks with exponentially distributed service times. The most commonly used decomposition method is single station decomposition, which dates back to the work of Hillier and Boling (1967), who considered tandem single server networks (Osorio and Bierlaire, 2007). The most common approach concerns single server feed-tandem single networks where each station is modeled as an M\M\1 station (Takahashi et al., 1980) but Koizumi et al. (2005) extended the approach to M/M/c stations. Here, each station is modeled as M\M\c queue for which closed form expressions of the performance measure exit separately. Most of the recent studies, on application of refined decomposition methods to multiple sever networks were conducted in advanced countries and there are few or no researchers using decomposition methods to multiple server networks involving both infinite and finite queue capacity on healthcare issues in Nigeria. Reason has been that resources allocations in Nigerian health-care centres is considered as being too complicated, thus, some rule-of-thumb approaches are employed despite the usefulness and relevance of operations research based model.

Materials and Methods

The Case Hospital and Problem Description

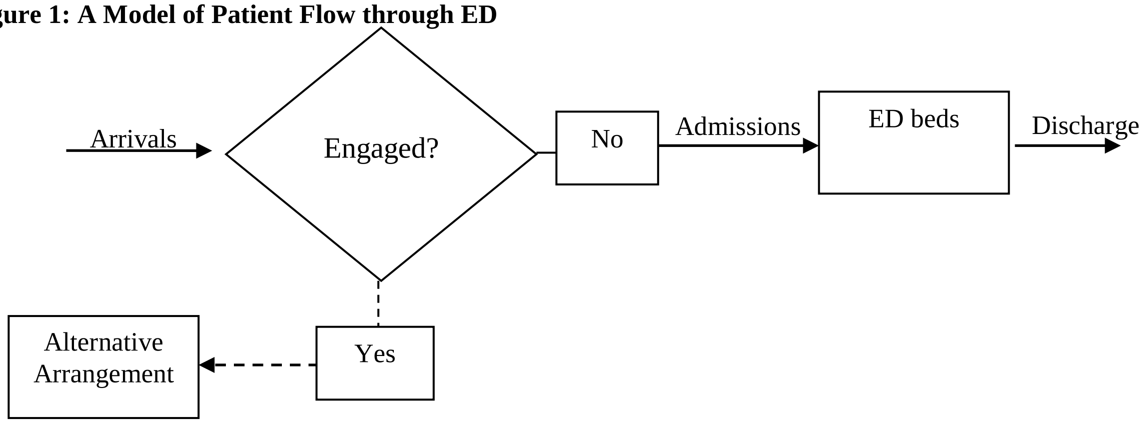
Lagos University Teaching Hospital, located in Lagos State, the most populous and highly competitive environment in Nigeria. The hospital is one of the best and largest Teaching/Specialist Hospitals in the country with a high concentration of skilled medical and paramedical staff in different areas of medicine. It has a successful transition through major changes in healthcare environment in Nigeria. The ED of University of Lagos Teaching Hospital is one of the busiest in Nigeria, serving about 50, 000 patients annually. The hospital faces a heavy patient load over emergency department (ED) beds as compared to general beds. Patient flow in the LUTH ED is very intense, as a result, overcrowding and delays are the major problems in the departments. Like every ED overall world, the management of LUTH has been working hard towards improving patient flow and reducing waiting time's in consonance with the National Health Service (NHS) target of "no-one should be waiting more than four hours in Accident and Emergency from arrival to admission, transfer or discharge. Average waiting times in accident and emergency will fall as a result to 75 minutes. ... …. … if they (patients) need a hospital bed they should be admitted to one without undue delay" (Smith and Mayhew, 2007). The researcher was of the opinion that effective allocation of the most important resource (bed) in this unit would increase the capacity-utilization of the hospital in terms of patient flow because its tasks and operations affect nearly all departments of the hospital. ED's performance in terms of patients flow and of the available resources can be studied using the queuing theory (Vass and Szabo, 2015)

Model Description

This study adopted queuing models in form of M/M/∞ and M/M/C/C/FCFS. Where, the first and the second M indicate a Markovian arrival and service times respectively. The ∞ indicates infinity servers and first C represents multiple servers (C>1), which implies that the capacity of the system is limited to the number of servers. The second C measures the capacity of the system in terms limited number of resources (beds and waiting space). ∞ and C represent number of beds at ED. The ED was modeled as a multi-servers single-stage system of identical parallel servers that process randomly patterned arrivals according to exponentially distributed service times. Each ED bed was treated as one service facility (server). M/M/∞ model assumes infinite number of beds at ED, reflecting the fact that all patients are accommodated or served because of the availability of unlimited beds. That is, no referral is rejected or delayed (blocked). In is this study, this model was used to quantify the impact of fluctuations in arrivals and variation in length of stay on bed requirements. The expected result indicates the efficiency level of the unit. This is very relevant because one of the main goals of the hospital is providing an admission guarantee for all arriving patients (Arnoud et al, 2006).

Secondly, M/M/C/C model assumes poisson arrival, exponential service times, multiple but finite number of servers-limited beds (limited capacity). M/M/∞ has infinite servers (beds) which is not practical in nature. This necessitates our second model – M/M/C/C. The model has This is a two-dimensional model called delay model or lost model, where all the patients that arrive while the station is full are delayed or considered to be lost. Blocking is a crucial feature of health care system and is directly related to the quality of health care delivery. Ridge et al (1995) described non-linear relationship between number of beds, mean occupancy level and the number of patients that have been transferred through lack of bed space. The treatment decision at ED and the entire LUTH is based on a FCFS queuing discipline, except that the patients referred internally, occasionally receive priority over external patients (i.e. patients from outside the system) and patients with serious health challenges are always treated first. The structural model of patient flow through ED is shown in Figure 1.

Figure 1: A Model of Patient Flow through ED



Source: Obamiro (2010)

Parameters Estimation

Patient arrival at the clinics ED is recorded when it is referred by a doctor either internally and externally. The queuing models adopted assumes that daily admission rates (average arrivals) follow a Poisson distribution (coefficient of variation=1) in consonance with some studies which have found that the arrival rate of patients to EDs follows a poisson distribution (McManus et al, 2004: Green, 2002; Arnoud, et al, 2007; Kauhanen, Kulvik, Kulvik, Maijanen and Martikainen, 2013; Liu, Xie and Liu, 2014)

The obtained data show that the total number of arrival for 18 months fluctuates around 24,263 for LUTH ED, the average number of patients arriving per day is therefore 44.37 patients. The length of stay was calculated as discharge-admission date, which is characterized by a relatively high variability. The unit records revealed that some patients had not exited from the clinics, thereby creating truncated LOA data. The reasons for this appear may be medical officers in charge of documentation may be engaged at time of discharged of some patients, so the estimated LOS of patient in LUTH ED 1.5 days. The capacity of the system (ED) is determined as the sum of the number of beds at the unit and the maximum number of patients allowed waiting or queuing at the unit.

The research received approval from LUTH Review and Ethical Committee. The necessary data were gathered from database and hospital files containing patient information for 18 months. It is worth mentioning, in some cases, patients discharge times were not indicated, so we could not determine the average length of stay of those patients. To avoid complexity, those patients' records were excluded from this research. Thus, the exclusion has some implications on the average length of stay in the two units.

RESULTS AND DISCUSSION

Fluctuation in Arrival and Variation in Length of Stay and Bed Allocation

The EDs patient flow of the case hospital was analysed using TORA Optimization software. The TORA is a Windows-based software application for solving life applicable problems or situation that can be modeled using queuing theory, linear programming model, transportation model, etc. (Taha, 2003). The steady state analysis of ED using the M/M/∞ to model is presented in table 1. This implies that the queuing experiences birth-death situation. This model helps to determine the efficiency and effectiveness of the selected units.

Queuing Output Analysis

Table 1: LUTH Emergency Department Analysis

Scenario 1—(M/M/9999) (GD/infinity/infinity)

Lambda	= 44.27000		Mu	=	1.50000
Lambda eff	=	44.27000	Rho/c	=	0.00295
Ls	=	29.51333	Lq	=	0.00000
Ws	=	0.66667	Wq	=	0.00000

N	Probability, pn	Cumulative, Pn
0	0.00000	0.00000
1	0.00000	0.00000
2	0.00000	0.00000
3	0.00000	0.00000
4	0.00000	0.00000
5	0.00000	0.00000
6	0.00000	0.00000
7	0.00000	0.00000
8	0.00000	0.00000
9	0.00001	0.00001
10	0.00002	0.00003

n	Probability, pn	Cumulative, Pn
11	0.00006	0.00009
12	0.00014	0.00023
13	0.00032	0.00054
14	0.00066	0.00121
15	0.00131	0.00251
16	0.00241	0.00492
17	0.00419	0.00911
18	0.00686	0.01597
19	0.01066	0.02663
20	0.01573	0.04236
21	0.02211	0.06447

Source: 2015

Modeling the Emergency Department to Determine Optimal Bed

Table 2 presents the results of the analysis of LUTH ED using an Erlang's delay probability- a queuing multiple model with a fixed number of beds and limited space (M/M/C/C). The aim of this queuing model is to determine the actual number of beds that serve a given population using the performance parameters to determine the probability of patients that are delayed waiting for beds. One advantage of using this model is that given an arrival rate, an average duration, and the number of servers, closed form expression for performance measures such as the probability of a positive delay or the expected delay can be easily determined (Green and Nguyen, 200).

Using λ of 44.27 patients/hour and length of stay (ALOS) of 1.5 days and looking at the effect of unit size (bed available) and occupancy levels on delays, table 2 shows probability of delay (p_D) as a function of number of beds for utilization levels ranging from 50% to 70%. This utilization range is guided by the current average occupancy rate for nonprofit hospitals which is about 63%, (AHA, 1996). Probability of delay was used to estimate performance parameters of emergency unit who require beds immediately. Probability of delay is commonly used as the standard measurement in service systems where customers have low tolerance for any delays. The analysis reveals that if it is desirable to keep p_D below 0.01(1%), it can only be achieved in ED (LUTH) that operate more than 64 beds. Yet some teaching hospitals or smaller public hospitals, often in less populated areas, may have far fewer beds. On the other hand, LUTH which has 61 beds for both adults and pediatrics could increase its occupancy level higher than the current rate. But increasing occupancy rate leads to higher probability of delay, hence, higher waiting time. It is pertinent to note that to the best of our knowledge, there exists no operational standard regarding patient delays for ED beds but past studies on analyzing the need for emergency beds used a p_D target of 10% (Green and Ngugen, 2001) and .01 (Schneider, 1981). In this study, 0.01 (1%) was adopted because of the size of the hospital and its moderate utilization. Probability of delay ($p_{D)}$ target of 1% can only be used but mostly suitable for large unit with high utilization (Green and Ngugen, 2001). This information is clearly important for hospital decision makers for capacity planning i.e. optimal bed determination.

Table 2: Probability of Delay with existing Occupancy rate and size at ED of LUTH

Numbers of bed		PD
10	1 – 0.00003	0.999
15	1 – 0.00251	0.997
20	1 – 0.04236	0.957
25	1 – 0.23424	0.765
30	1 – 0.58366	0.416
35	1 – 0.86376	0.13
40	1 – 0.97374	0.02
45	1 – 0.99704	0.0029
50	1 – 0.99979	0.00021
55	1 – 0.99999	0.00005

Source: 2015

Conclusion and Recommendations

The 18-month experience illustrates that queuing network model may be used to allocate ED resource (bed) in a unit that operate at or near capacity. This study accurately determines the number of bed suitable for ED of the case hospital to serve the visiting patients in order to reduce waiting times, which invariably should lead to a better patient flow and improve service quality within the departments and entire the hospitals. Generally, improving patient flow, and reducing waiting time for patient can be achieved through the following ways: (1) managing the arrival rate by different appointment methods such as online appointment management system to spread arrivals and call in appointment system; (2) increasing the number of servers; (3) optimizing the service rate. It is worth mentioning that a good patient flow reduces waiting and facilitates bed usage, and optimizes patient length of stay

References

Kyoung, W. C., Seong, M. K., Young, M. C., Yong, U. S. (2017). Application of queuing theory to the analysis of changes in outpatients' waiting times in hospitals introducing EMR, Healthc Inform Res., 23(1); 35-42.

Vass, H. and Szabo, K. Z. (2015). Procedia Economics and Finance 32, 479 – 487

Nidhi Malik1, O. K. Belwal2 (2016). **A**pplication of queuing theory to patient satisfaction at combined hospital, Srinagar Garhwal Uttarakhand, International Journal of Science and Research (IJSR), 5 (4), 1621-1624.

Bahadori, M., Mohammadnejhad, S. M., Ravangard, R. Teymourzadeh, E. (2014). Using queuing theory and simulation model to optimize hospital Pharmacy performance. Iran Red Crescent Med J. 2014 March; 16(3):

Aboukanda, E. K. and Latif, M (2014). The effect of patient behaviour on wait times in emergency departments, international journal of business and commerce, 3(6): 18-31

Agrawal, G. (2010). A queuing model for health Centre, international electronics and information engineering, Indian institute of technology, Kanpur, 542-547.

Arnoud, M. d, Rossum, A.C., Visser, M.C. and Koole, G.M. (2006), Bottleneck analysis of emergency cardiac in-patient flow in a university: an application of queuing theory. *Invest Med*, 28, (6): 316-317

Asplin, R. B., Magid, J. D., Rhodes, V. K., Solberg, I. L., Lurie, N., and Camargo, A. C. (2003). A conceptual model of emergency department crowding, annals of emergency medicare 42 (2): 173-180.

Au-Yeung SWM, Harder U, McCoy EJ, (2009). Predicting patient arrivals to an accident and emergency department. *Emerg Med J ;**26**:241–4.*

Av-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.

Bruin, A. M., Rossum, A. C., Visser, M. C. and Koole, G. M. (2007). Modeling the emergency cardiac in-patient flow: an application of queuing theory, health care management science, 10, 125-137.

Creemers, S. And Lambrecht, M.R. (2007). Modeling a healthcare system as a queuing Network: the case of Belzian hospital. A paper submitted to department of decision sciences and information management, research centre for operations and management, Katholieke university Leuven.

Delia, D. (2007). Hospital capacity, patient flow and emergency department use in New Jersey, a report to the new jersey department of health and senior services.

Derlet, R. & Richards. J. (2000). Overcrowding in the nation's emergency departments: complex causes and distributing effects, annual of emergency Medicare, 35(1): 63-68.

Diakogiannis, A. (2014). Improving the performance of emergency departments: A survey from an operations management perspective, Laboratoile genieiIndustriel

Gauld, R, (2000). Solving the surgical waiting list problem: New Zealand's booking system, int. journal of health planning management , 15,(4): 257-272.

Georgievskiy, I., Georgievskiy, Z. and Pinney, W. (2007). Using computer simulation modeling to reduce waiting times in emergency departments, Springer, First Edition, pp 1-18**.**

Geranmayeh, S., and Iyer, A. S. (2008). capacity analysis of critical hospital resources using system dynamics approach, Unpublished Manuscript.

Gorunescu, F. Mcleen S.I, Millard, P. H. (2002). Using a queuing model to help plan bed allocation in a department of geriatric medicine. health care management science, (5): 307 – 312.

Green, L.V., (2002). How many hospital beds? Inquiry, 37 (4), 400-412.

Green, V.L., Soares, J., Giulio, J. and Green, R.(2006). Using queuing theory to increase the effectiveness of emergency department provides staffing, academic emergency medicine, 13, (1): 61-68.

Gross, D. and Harris, .C.M (1998). Fundamentals of queuing theory, New York: Wiley, 3rd Edition.

Hall, R., Benson, D., Mural, P. and Dessouky, M., (2001). Modeling patent flows through the healthcare system, Revised Manuscript.

Hall, W.R. (2006). Patient flow: the new queuing theory for healthcare, OR/MS, Today.

Henry, M. C. (2001). Overcrowding in America's emergency department inpatient wards replace emergency Car, academic emergency medicine, 8 (2): 151 – 155.

Hiller, F.S. and Boling, R.W. (1967) Finite queues in series with exponential or Erlang service times- a numerical approach. Operations Research**,** 15: 286-303

Hiller, S.F and Lieberman, J.G. (2005). Introduction to operations *research*. Boston: McGraw Hill, Eight Edition.

Institute of Medicine (IOM) (2006). Hospital-based emergency care: at the breaking point, national academy of sciences: Washington DC.

Jackson, J.K (1957). Networks of waiting lines, operations research, 5: 518-522

Jackson, J.K (1963). Job shop like queuing systems, management science, 10: 131-142.

Koizumi, N. (2002). A queuing analysis with blocking: analysis of congested patients flows in mental health systems, regional science, university of Pennsylvania,

Koizumi, N., Kuno E and Smith, E.T (2005). A queuing network model with blocking: analysis of congested patient flow in mental health systems, health care management science, *8(1):49-60*

LewinGroup (2004). Hospital capacity and emergency department diversion: four community case studies, American health association.

Liu, Q., Xie, X. and Liu, R. (2014). A queuing approximation method for capacity planning of emergency department with time-varying demand, ieee international conference on automation science and engineering (CASE) Taipei, Taiwan, August, 558-563.

McManus, M.C, Long, M.C, Cooper A.B and Litvak, E(2004). Queuing theory accurately models the need for critical care resources, Anesthesiology, 100, (5); 1271-1276.

Medhi, J.(2003). Stochastic models in queuing theory. Amsterdam: Academic Press, Second Edition.

Obamiro, J. K. (2012). Queuing patient flow and resource allocation in intensive care unit, Bayero business review, 4(2), 39-50 (University of Bayero, Kano State).

Obamiro, J. K. (2013). Effects of waiting time on patient satisfaction: Nigerian hospital experience, journal of economic behaviour, journal of economic behaviour, 3(3): 7-17.

Olorunsola S. A, Adeleke R. A and Ogunlade T. O (2014). queueing analysis of patient flow in hospital, IOSR journal of mathematics (IOSR-JM), 10 (4):47-53.

Osorio and Bielaire (2007). Describing network congestion with an analytic queuing to network model.,Swiss Transport Research Conference

Ridge, J., Jones, S Nielsen,M. and Shahani, A.,(1998). Capacity Planning for Intensive Care Units, European Journal of Operational Research, 105:346-355

Shmueli, a., Springe, C. and Kaplan, E.,(2003). Optimizing admissions to an intensive care unit, health care management science, 6:113-136

Smith, D., and Mayhew, L., (2008). Using queuing theory to analyse governments 4 –h completion time target in accident and emergency departments, health care management science, 11: 11 -21.

Taha A.H. (2002) Operations Research: An Introduction, Delhi: Pearson Prentice Hall, Seventh Edition.

Takahashi, Y., Miyahara, H. and Hasegewa, T., (1980). An Approximation Method for Open Restricted Queuing Networks. Operations Research, 28(3): 594-602.

Vandale, N., De Boeck, L. & Callewier, D. (2002). An Open Queuing Network for Lead Time Analysis. IIE Transaction, 34, pp 1- 9.

Vasanawala, S.S. and Desser, T.S. (2005). Accommodation of Requests for Emergency US and CT: Application of Queuing Theory to Scheduling of Urgent Studies. Radiology, 238, (1): 244-249.

Xie, H., Chaussalet, T. And Pees, M. (2007). A Semi-open Queuing Network Approach to the Analysis of Patient Flow in Healthcare Sstem. Twentieth IEEE International Symposium on Computer Based Medical Systems, 0-7695-2905-4/07.

Armony, M., Israelit, S., Mandebaum, A., Marmor, N. Y., Tseytlin, Y. and Yom-to, B. G. (2015) On patient flow in hospitals: a data-based queueing-science perspective,Stochastic Systems, 5(1): 146-194

Volira, S., Dulta, G., and Ghosh, D. K. (2008) "Capacity Management of Intensive Care Units in a Multi-specialty Hospital in India", *Indian Institute of Management Ahmedabad, Research and Publication Department Series IIMA Working Papers , number WP2006-07-04*.

Caglar, T. (2003) "A Queuing Theoretic Approach to Gridlock prediction in Emergency Department". *An M.Sc. Project Submitted to the Department of Industrial and Systems Engineering.*

Kauhanen, A., Kulvik, M., Kulvik, S. Maijanen S. and Martikainen, S. (2013) ETLA
Working Papers No. 10 ,23