



Performance Measure of Queue Model Technology in Health Sector

Bello Muriana¹; Ogba Paul²

¹Information Technology and Resource Centre, Kogi State University, Anyigba, Nigeria

²Information Technology and Resource Centre, Kogi State University, Anyigba, Nigeria

¹muriana.b@ksu.edu.ng; ²Ogba.p@ksu.edu.ng

Abstract- Patients who wish to seek medical care often appear at random, and at the moment they need urgent treatment. When the service facilities work at full capacity upon arrival they would have to wait. At this stage, a queue, depending on the difference between the arrival rate of the patients, time needed for services and the time patients wait to be served is formed. This paper analyzes how queue model can be used to achieve an optimal balance between the service cost and the waiting cost, its performance measure with the help of Measurement scales for a queuing system such as, the average number of customers in the queue (L_q), the average number of customers in the system including the entity being served (L_s), the average waiting time in the queue (W_q), and the average waiting time in the system (W_s).

Keywords - Medical Sector, Queue, Performance Measure, Measuring Scale, Waiting Time, Arrival Time.

I. Introduction

Many medical-care facilities face lengthy waiting time. In any hospital long waiting time is perceived a sign of poor quality that needs improvement. Medical-care can be enhanced by understanding how the complexities of the queues function in real life. The key aim of medical-care management is to satisfy patients effectively, consistently and efficiently by offering less waiting times and delays with increased employee satisfaction. There are long queues at hospitals in the cities thus; patients have to wait for months to get even small operations due to having enough staff and inefficient usage of resources. This study focuses on queues models that can be used to decrease waiting time. In medical-care institutions, the adverse effect of queuing in relation to time spent in a queue for patients to access treatments is increasingly becoming a

huge source of concern for modern society that currently exposed to significant strides in technological advancement and the danger of keeping customers waiting could become a cost to them. The cost of waiting for space, facilities, equipment, and supplies depend on the patient waiting time. Using the estimation of the waiting cost helps decision makers to have the ability to evaluate the recommended number of servers by minimizing the total cost including the service cost and the waiting time. Queuing models can be very useful in identifying appropriate levels of staff, equipment, and beds as well as in making decisions about resource allocation and the design of new services. Queuing models require very little data and result in relatively simple formulae for predicting various performance measures such as mean delay or probability of waiting more than a given amount of time before being served. This means that they are easier and cheaper to use and can be more readily used to find optimal solutions rather than just estimating the system performance for a given scenario.

II. Related Work

Many organizations routinely use queuing models to help determine capacity levels needed to respond to experienced demands in a timely fashion. The queue discipline refers to the order in which members of the queue are selected for service (Hillier F., Lieberman G). The usual queue discipline is first come, first served (FCFS), where customers are served in order of arrival. Improving quality and safety and reducing medical-care costs have become more important goals in an age of healthcare reform than ever before (Singh, 2007). Service quality is affected not only by the actual waiting time but also by the perceived waiting time. The act of waiting time has significantly influenced customers' view (Obamiro, 2010). More so, researchers, over time have found out that customers' perception is affected not just by waiting time but also by customer expectations and view of the causes for the waiting (Taylor, 1994). Invariably, the critical issues in queue management are not only the actual amount of time the customer has to wait, but also the customer's views of that wait (Davis and Heineke, 1994). Medical-care systems operation such as patient scheduling, resource scheduling, queue length, Limited Queue Discipline (LQD), blocking, medical-care systems design and analysis are crucial parts of study in healthcare management. The effective usage of resources, high quality of service and less queues are the main aims that needed to be evaluated for the successes of clinics and hospitals (Lakshmi & Sivakumar, 2013).

III. Model Specification

This study adopted M/M/C: (FCFS/ ∞/∞). This model assumes a single queue with unlimited waiting room that feeds into multiple identical servers. It is assumed that the arrivals follow a Poisson distribution at an average of λ patients (pregnant women). The service times are exponentially distributed, with an average of μ patients (pregnant women) per unit of time and number of servers C. with infinite system limit (∞) and total source limit (∞).

If there are n pregnant women in the queuing system at any point in time, then the following two cases may arise:

1. If $n < C$, the number of pregnant women in the system is less than the number of servers, then there will be no queue.
2. If $n \geq c$, the number of pregnant women in the system exceeds or equivalent to the number of servers (doctors) then all servers will be busy.

For the purpose of modeling, the arrivals (n) are the pregnant women. The doctors are the servers(C). Medhi (2003) suggests that whenever there are ($n \geq C$) in the system, then all the C channels are busy. The following system performance parameters were employed in the study and defined as follows;

$$\rho = \lambda / C\mu \quad (1)$$

$$L_q = \frac{\rho^2}{(1-\rho)} \quad (2)$$

$$W_q = \frac{L_q}{\lambda} \quad (3)$$

$$W_s = W_q + \frac{1}{\mu} \quad (4)$$

$$L_s = \lambda W_s \quad (5)$$

$$P_0 = 1 - \rho \quad (6)$$

λ = arrival rate of patient per unit time

μ = service rate (length of stay) of a patient (expectant woman) per unit time;

C = Number of servers (doctors) in the model

ρ = system utilization factor.

W_q = average waiting time of an patients in the queue

L_q = average number of patients in the queue

W_s = average waiting time of patients in the system

L_s = average number of patients in the system

IV. Cost Model

In order to determine and identify the optimal number of servers in the system, it is important to weigh two important costs when making these decisions; service cost and waiting cost.

Therefore, the total cost optimization problem can be expressed as;

$$EWC = \lambda W_s \times C_w \quad (7)$$

$$ETC(c) = EOC(c) + EWC(c) \quad (8)$$

$$= C_o + CC_s + \lambda W_s (C_w) \quad (9)$$

Where ;

c = number of servers

C_o = the fixed cost of operation system per hour

C_s = service cost of each server

C_w = Expected waiting cost of patients

EWC = Expected waiting cost

ETC = Expected Total Cost per unit time,

V. Method

This study adopts a direct observations study approach. In-depth study of Federal Medical Center (FMC), Ido Obstetrics' unit, attendance reports for three weeks were made. Interviews with the management, Doctors, and records of staffs were conducted to validate the secondary data and to accumulate other relevant information that was used to model patients' (pregnant women) arrival and length of stay. In this study 100 pregnant women were interviewed, the data were obtained

from obstetrics unit of FMC Ido, on allocated weeks for the obstetrics outpatients' visit to the teaching Hospital. A systematic sampling approach was used where a stratum is reflected every week. The rate at which pregnant women come for a visit per hour and also the service rate per hour also observed and recorded. The following assumptions were made;

1. The arrival of patients (expectant mothers) follows a Poisson Probability distribution at an average rate of λ pregnant women per hour.
2. The queue discipline is First-Come, First-Served (FCFS). Before any of the server there is no particular preference to any arrival.
3. Service times are distributed exponentially, with an average of μ patients per hour.
4. The average arrival rate is greater than the mean service rate.
5. Server, in this case, depicts only Doctors.

VI. Results and Discussion

The following input data was used to evaluate the performance measure of the multi-server queue system at FMC Obstetrics' unit.

Table 1: input parameters

Parameter M/M/C:FCFS/ ∞/∞	Value
Arrival rate(λ)	8 Pregnant women per hour
Service rate(μ)	3 Pregnant women per hour
Number of Servers	3,4 5,6,7

Table 2: performance measures of M/M/C/FCFS/ ∞/∞

	C	λ	μ	P_0	L_s	L_q	W_s	W_q	ρ
1	3	8	3	0.11	9.05	6.38	1.13	0.79	0.89
2	4	8	3	0.33	3.42	0.76	0.43	0.09	0.67
3	5	8	3	0.47	2.85	0.18	0.36	0.02	0.53
4	6	8	3	0.56	2.72	0.05	0.34	0.01	0.44
5	7	8	3	0.62	2.68	0.01	0.34	0.00	0.38

Table 3: Summary of the Marginal Costs, Service Costs, Opportunity Cost and the Number of Servers

	C	λ	W_s	λW_s	$C_o(\#)$	$C_s(\#)$	$C_w(\#)$	$CC_s(\#)$	$\lambda W_s C_w(\#)$	ETC(#)
1	3	8	1.131	9.05	1200.6	4550.2	9020.5	13650.6	81635.5	96486.7
2	4	8	0.428	3.42	1200.6	4550.2	9020.5	18200.8	30850.1	50251.5
3	5	8	0.356	2.85	1200.6	4550.2	9020.5	22752.5	25708.4	49661.5
4	6	8	0.340	2.72	1200.6	4550.2	9020.5	27301.2	24535.8	53037.6
5	7	8	0.335	2.68	1200.6	4550.2	9020.5	31851.4	24174.9	57226.9

Table 4: Summary analysis of the queuing model

Performance Measure	3 Servers	4 Servers	5 Servers	6 Servers	7 Servers
Arrival rate (λ)/h	8	8	8	8	8
Service rate (μ)/h	3	3	3	3	3
Utilization (ρ)	88.9%	66.7%	53.3%	44.4%	38.1%
L_s/h	9.0467	3.4235	2.8514	2.7162	2.6798
L_q/h	6.3801	0.7568	0.1847	0.0496	0.0131
W_s/h	1.1308	0.4279	0.3564	0.3395	0.0016
W_q/h	0.7975	0.0946	0.0231	0.0062	0.0016
P_0	0.111	0.333	0.467	0.556	0.619
TC(#)	96486.7	50251.5	49661.5	53037.6	57226.9

Figure 1 depicts the line plots of the Utilization factor (ρ) against Average patients' waiting time in the system. Figure 2 shows average Number of patients in the system against the probability of the system is idle and figure 3 shows the expected Total Cost against the Server.

Figure 1: Utilisation Factor (ρ) against Average Pregnant Women waiting time in the System (W_s)

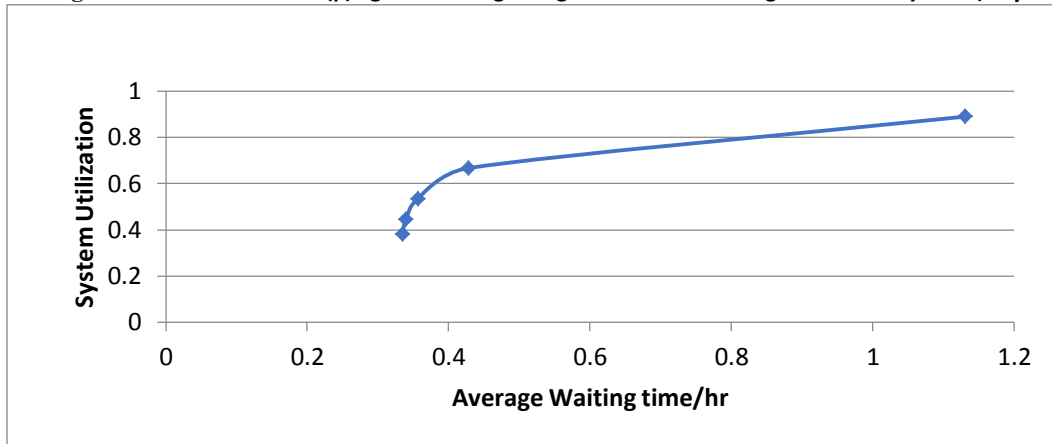


Figure 2: Average Number of Pregnant Women in the System (L_s) against Probability of the System being idle (P_o)

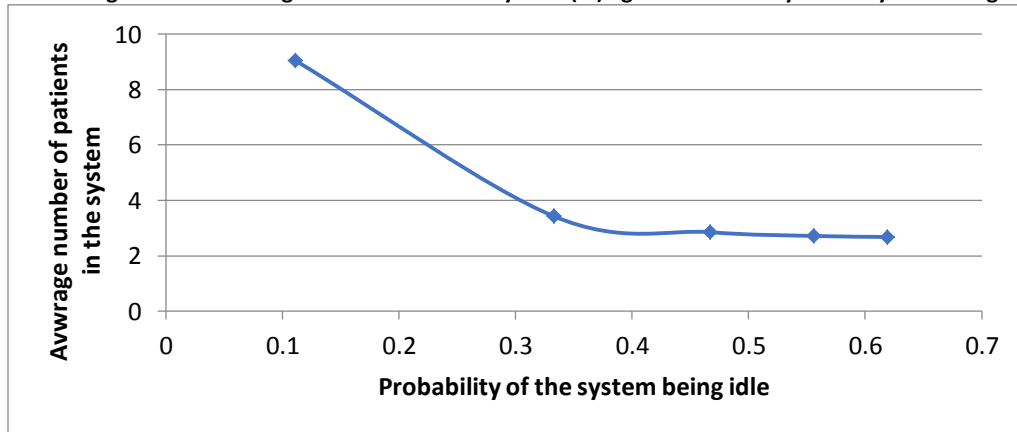
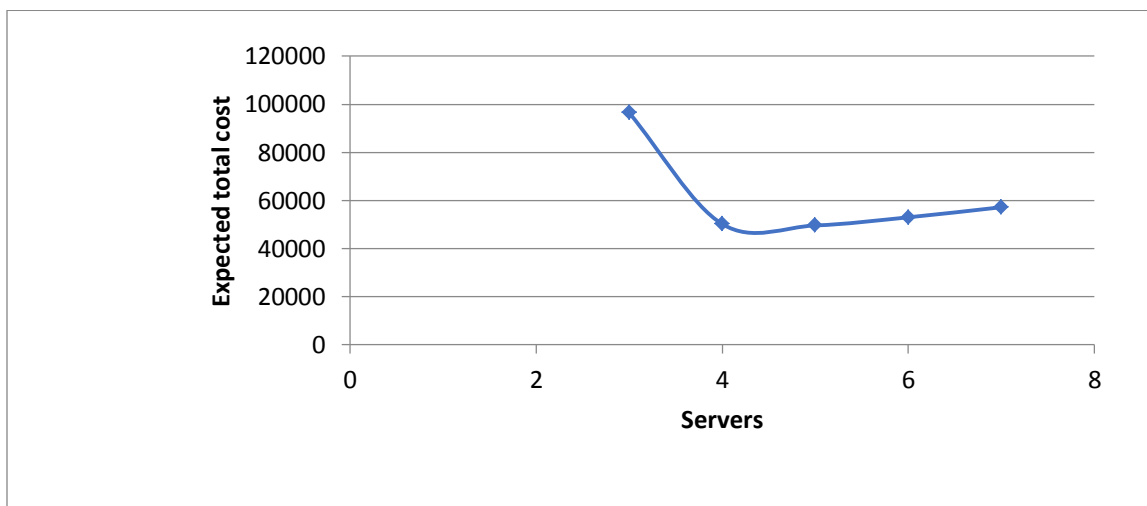


Figure 3: Expected Total Cost against Servers (Doctors)



The Figures 1, 2 and 3 above; depicts the extremely marginal changes for scenario 1, 2 and 3 on the average waiting time, average time spends in the system, the length of a queue, the number of patients in the system, system utilization factor, the probability of no patient in the queue (P_0) and the total system cost. They all work with the empirical graphs above but one must be vigilant about the costs involved in bringing in such modest changes. Sustainability requires an excellent, outstanding balance between the number of doctors, costs and optimal system performance. From the graph depicted above, it would mean taking on additional cost to hire more doctors.

Hence, the conclusion that scenario 3 with five Doctors at the post is best for optimal performance with a total system cost of #49661.50. It is also worth noting that the patients' waiting time, cost and congestion in the system is less at this optimal server level.

Summary

Established Knowledge of the use of queuing model to evaluate program parameters is of the utmost importance for healthcare providers trying to attract, retain and provide a patient with quality healthcare with the ever-competitive "global market." Queuing theory is a well-known and validated statistical approach to the study of waiting lines. The queuing characteristics at FMC Obstetrics' were evaluated using a multi-server queuing model and the costs of waiting and service received to assess the optimum level of service. The empirical findings showed the average length of line, waiting time for patients as well as over-use of doctors, when the ability of doctors in the clinic is expanded from three to five, the overall cost of waiting and service costs can be minimized. The operation directors should understand the trade-off that must take place between the expense of delivering outstanding service and the benefit of waiting time for customers. Service improves and the cost of time spent waiting in the line decreases.

References

1. Adedayo, O. A., O., Obamiro, J. K.,(2006)Operations Research in Decision Analysis and Production Management. Lagos: Pugmark Nig.Ltd.
2. Ben-Tovim, D.I., Bassham, J.E. and Bennett, D.M. 2008. Redesigning care at the Flinders Medical Centre: clinical process redesign using "lean thinking", *Med J Aust*,188 (5 Suppl), S27-S31.
3. Davis, M.M., Heineke, J.,(1998),How Disconfirmation, Perception and Actual Waiting Times Impact, Unpublished Manuscript.

4. Fomundam, S. and Herrmann, J.W. 2007. A survey of queuing theory applications in healthcare. ISR technical report, Technical Report 2007-24, College Park (MD): Institute for Systems Research, University of Maryland.
5. Hillier F., Lieberman G. Introduction to Operations Research. 7th ed. New York: McGraw Hill; 2001.
6. Obamiro, J.K.(2003), Queuing theory and Patient Satisfaction, An overview of Terminology & Application in Ante-natal Care Unit.<http://www.upg-bulletin-se.ro>.
7. Singh, V. “Use of Queuing Models in Health Care”, Department of Health Policy and Management, University of Arkansas for medical science 2007.
8. Strecher, V.J., De Vellis, B.M., Becker, M.H. and Rosenstock, I.M. 1986. The role of self-efficacy in achieving health behavior change, *Health Educ Q*,13, 73-92.
9. Taylor, S.,(1994), Waiting for Service: The Relationship between Delays and Evaluation of Service, *Journal of Marketing*, 58(2): 56-6.