A Study of Waiting And Service Costs of A Multi-Server Queuing Model In A Specialist Hospital

Kembe, M. M, Onah, E. S, Iorkegh, S

Abstract:- The effect of queuing in relation to the time spent by patients to access clinical services is increasingly becoming a major source of concern to most health –care providers. This is because keeping patients waiting too long could result to cost to them (waiting cost). Providing too much service capacity to operate a system involves excessive cost. But not providing enough service capacity results in excessive waiting time and cost. In this study, the queuing characteristics at the Riverside Specialist Clinic of the Federal Medical Centre , Makurdi was analysed using a Multi-server queuing Model and the Waiting and service Costs determined with a view to determining the optimal service level. Data for this study was collected at the Riverside specialist clinic for four weeks through observations, interviews and by administering questionnaire. The data was analysed using TORA optimization Software as well as using descriptive analysis. The results of the analysis showed that average queue length, waiting time of patients as well as overutilization of doctors at the Clinic could be reduced at an optimal server level of 12 doctors and at a minimum total cost as against the present server level of 10 doctors with high Total Cost which include waiting and service costs. This model can also be used by decision and other policy makers to solve other Multi-server queuing problems.

Keywords:- Service Cost, Servers, Utilization factor, Waiting Cost

1. Introduction

A common situation that occurs in everyday life is that of queuing or waiting in line .Queues (waiting lines) are usually seen at bus stops, hospitals, bank counters and so on [15]. In general, queues form when the demand for service exceeds its supply [5]. Wait time depends on the number of customers (human being or objects) on queue, the number of servers serving line, and the amount of service time for each individual customers. In healthcare institutions the effect of queuing in relation to the time spent for patients to access treatment is increasingly becoming a major source of concern to a modern society that is currently exposes to great strides in technological advancement and speed [16]. The danger of keeping customers waiting could become a cost to them [3]. The time wasted on the queue would have been judiciously utilised elsewhere (opportunity cost of time spent in queuing). In a waiting line system, managers must decide what level of service to offer. A low level of service may be inexpensive, at least in the short run, but may incur high costs of customer dissatisfaction, such as lost of future business. A high level of service will cost more to provide and will result in lower dissatisfaction costs. When considering improvements in services, the health care manager weighs the cost of providing a given level of service against the potential costs from having patients waiting .The goal of queuing is therefore to minimize the total cost to the system.

- Kembe, M.M., Visiting Lecturer at the University of Agriculture Makurdi, +2348036177129, kdzever@yahoo.com
- Onah, E.S., Professor of Mathematics at the University of Agriculture Makurdi
- Iorkegh,S., A Postgraduate Student, Department of Mathematics / Statistics / Computer Science, University of Agriculture, Makurdi

The two basic costs mentioned are costs associated with patients or customers having to wait for service (Wait Cost) which include loss of business as some patients might not be willing to wait for service and may decide to go to the competing organizations, cost due to delay in care or the value of the patients time (opportunity cost of the time spent in queuing) and decreased patients satisfaction and quality of care. While Service Cost is the cost of providing service .These includes salaries paid to employees, salaries paid to employees or servers while they wait for service from other servers [17]. Cost of waiting space, facilities, equipment, and supplies. Using the estimation of waiting cost allows decision makers to have the capability of determining the optimal number of servers by minimizing the total cost including the service cost and the waiting cost. The cost of waiting for every individual differs depending on what the individual earns every hour. Some might have their cost of waiting in multiples of other people's value. Queues or waiting lines or queuing theory, was first analyzed by A.K. Erlang a Danish Engineer in 1913 in the context of telephone facilities. He was experimenting with the fluctuating demand for telephone facilities and its effect on automatic dialling equipment at the Copenhagen telephone System. Since World War II this theory has been applied to many business and human service fields. Literature on queuing indicates that waiting in line or queue causes inconvenience to economic costs to individuals and organizations. Healthcare, airline companies, banks, manufacturing firms etc., try to minimize the total waiting cost, and the cost of providing service to their customers.

[8] Reviews research on models for evaluating the impact of bed assignment policies on utilisation, waiting time, and the probability of turning away patients.[10] reviewed the use of queuing theory in pharmacy application with particular attention to improving customer satisfaction. Customer satisfaction is improved by predicting and reducing waiting times and adjusting staffing. [18] 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 increase cost equal the benefits. [14] Considered a pharmacy 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. [4] Chose 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. The author also calculated 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 queue do not exceed a given threshold. [6] Incorporated advertising into their model to control the demand for laboratory services. The model assumes that clients would leave without service if they wait above a certain amount of time. Long waiting time of HIV-Aids patients and overutilization of medical personnel have been the major challenges facing the Riverside Specialist Clinic of the Federal Medical Centre Makurdi. In this study, Waiting and Service Costs at the clinic using a Multi-server queuing model with a view determining the optimal service level was studied.

2. Methods

Data for this study were collected from Riverside Specialist Clinic of federal Medical Centre Makurdi. The methods employed during data collection were direct observation and personal interview and questionnaire administering by the researcher. Data were collected for (4) weeks. The following assumptions were made for queuing system at the Riverside Specialist clinic which is in accordance with the queue theory. They are:

1. Arrivals follow a Poisson probability distribution at an average rate of λ customers (patients) per unit of time.
2. The queue discipline is First-Come, First-Served (FCFS) basis by any of the servers. There is no priority classification for any arrival.
3. Service times are distributed exponentially, with an average of μ patients per unit of time.
4. There is no limit to the number of the queue (infinite).
5. The service providers are working at their full capacity.
6. The average arrival rate is greater than average service rate.
7. Servers here represent only doctors but not other medical personnels.
8. Service rate is independent of line length; service providers do not go faster because the line is longer.

2.1 The M/M/S Model

The model adopted in this work is the (M/M/S) :(∞ /FCFS) - Multi-server Queuing Model .For this queuing system, it is assumed that the arrivals follow a Poisson probability distribution at an average of λ customers (patients) per unit of time. It is also assumed that they are served on a first-come, first-served basis by any of the servers (in these case doctors). The service times are distributed exponentially, with an average of μ customers (patients) per unit of time and number of servers S. If there are n customers in the queuing system at any point in time, then the following two cases may arise:

(i) If n<S,(number of customers in the system is less than the number of servers), then there will be no queue. However, (S–n) number of servers will not be busy. The combined service rate will then be $\mu_n = n\mu$; n<s

(ii) If $n \geq s$, (number of customers in the system is more than or equal to the number of servers) then all servers will be busy and the maximum number of customers in the queue will be (n – s). The combined service rate will be $\mu_n = s\mu$; n≥s

From the model the probability of having n customers in the system is given by

$$p_n = \begin{cases} \left(\rho^n/_{n!}\right) p_0 & n \leq s \\ \rho^n/_{(s!\, s^{n-s})}\, p_0 & n > s;\ \rho = \lambda/_{s\mu} \end{cases}$$

$$p_o = \left[\sum_{n=0}^{s-1} \frac{1}{n!}\left(\frac{\lambda}{\mu}\right)^n + \frac{1}{s!}\left(\frac{\lambda}{\mu}\right)^s \frac{s\mu}{s\mu-\lambda}\right]^{-1} \quad (1)$$

We now proceed to compute the performance measures of the queuing system.

The expected number of the customer (patients) waiting on the queue (length of line) is given as:

$$Lq = \left[\frac{1}{(s-1)!}\left(\frac{\lambda}{\mu}\right)^s \frac{\mu\lambda}{(\mu s-\lambda)^2}\right] p_o \quad (2)$$

Expected number of customers (patients) in the system

$$Ls = Lq + \frac{\lambda}{\mu} \quad (3)$$

Expected waiting time of customer (patients)in the queue

$$Wq = \frac{Lq}{\lambda} \quad (4)$$

Average time a customer (patient) spends in the system:

$$Ws = \frac{Ls}{\lambda} \quad (5)$$

Utilisation factor i.e the fraction of time servers (doctors) are busy

$$\rho = \frac{\lambda}{\mu s} \quad (6)$$

Where, λ = the arrival rate of patients per unit time, μ = the service rate per unit time, s = the number of servers, ρ_0 = the probability that there are no customers (patients) in the system, Lq = Expected number of customers in the queue, Ls = Expected number of customers in the system, Wq = Expected time a customer (patient) spends in the queue, Ws = Expected time a customer (patient) spend in the system.

2.2 Introducing Costs into the Model

In order to evaluate and determine the optimum number of servers in the system, two opposing costs must be considered in making these decisions: (i) Service costs (ii) Waiting time costs of customers. Economic analysis of these costs helps the management to make a trade-off between the increased costs of providing better service and

the decreased waiting time costs of customers derived from providing that service.

Expected Service Cost E (SC) = SC_S (7)

Where, S= number of servers, C_S= service cost of each server

Expected Waiting Costs in the System E (WC) = $(\lambda W_S)C_w$ (8)

Where λ =number of arrivals, Ws= Average time an arrival spends in the system

C_w= Opportunity cost of waiting by customers (patients)
Adding (7) and (8) we have,

Expected Total Costs E (TC) = E (SC) + E (WC) (9)

Expected Total Costs E (TC) = $SC_S + (\lambda Ws)\, C_w$ (10)

3 Analysis of Data

We use TORA software to compute the performance measures of the multi-server queuing system at the Riverside Specialist hospital using arrival rate (λ) =52 patients/hr, Service rate (μ) = 6 patients/hr and number of servers (S) = 10.

Table1. Performance Measures of Multiserver Queuing Model at the Riverside Specialist Hospital

Scenario	S	Lambda	Mu	L'da eff	p_o	Ls	Lq	Ws	Wq
1	10	52.00000	6.00000	52.00000	0.00012	12.39855	3.73188	0.23843	0.07177
2	11	52.00000	6.00000	52.00000	0.00015	10.00902	1.34235	0.19248	0.02581
3	12	52.00000	6.00000	52.00000	0.00016	9.23330	0.56663	0.17756	0.01090
4	13	52.00000	6.00000	52.00000	0.00017	8.91794	0.25128	0.17150	0.00483
5	14	52.00000	6.00000	52.00000	0.00017	8.77902	0.11235	0.16883	0.00216

Table 2: Summary analysis of the Multi-Server queuing Model of the Riverside Specialist Hospital

Performance Measure	10 Doctors	11 Doctors	12 Doctors	13 Doctors	14 Doctors
Arrival rate (λ)	52	52	52	52	52
Service rate(μ)	6	6	6	6	6
System Utilisation	86.6%	78.8%	**72.2%**	66.7%	61.9%
Ls	12.399	10.009	**9.233**	8.918	8.779
Lq	3.732	1.342	**0.567**	0.251	0.112
Ws – in hours	0.238	0.192	**0.178**	0.172	0.169
Wq – in hours	0.072	0.026	**0.011**	0.005	0.002
Po	0.012%	0.015%	**0.016%**	0.017%	0.017%
Total System Cost/hr	₦14,509.08	₦13,235.89	**₦13,174.84**	₦13,459.29	₦13,875.99

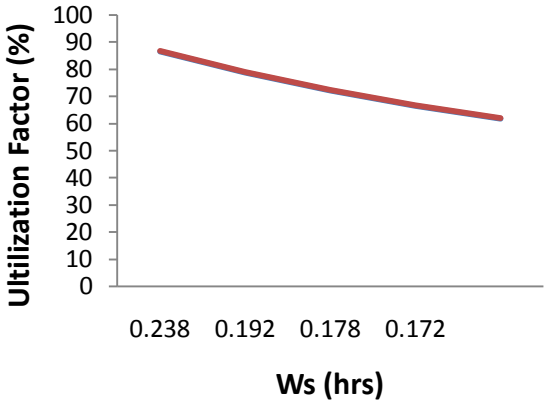

Fig.1: Ultilization Factor(ρ) against Average Patients Waiting time in the System (Ws)

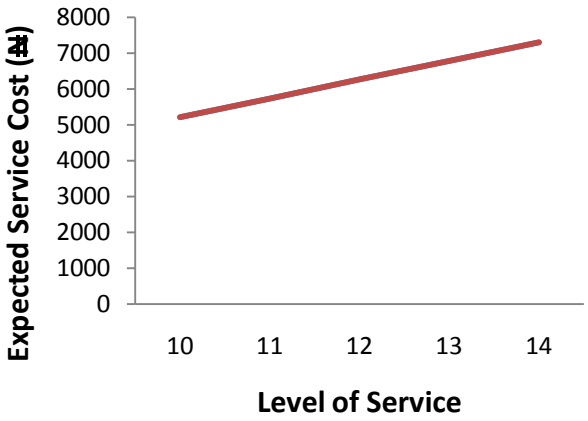

Fig.4: Expected Service Cost against Level of Service.

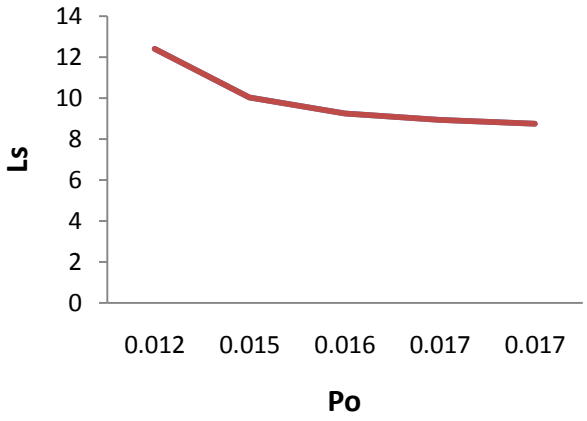

Fig. 2: Average Number of Patients in the System (Ls) against Probability of the system being idle(Po)

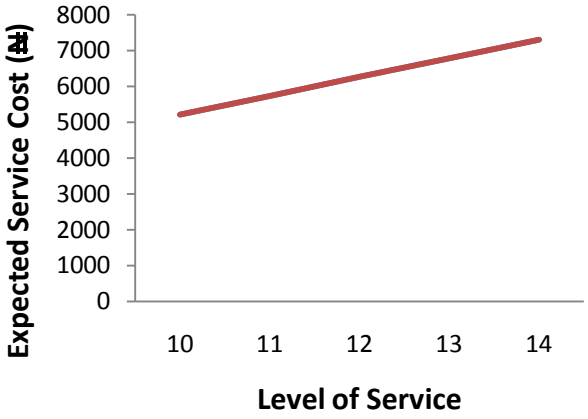

Fig.4: Expected Service Cost against Level of Service.

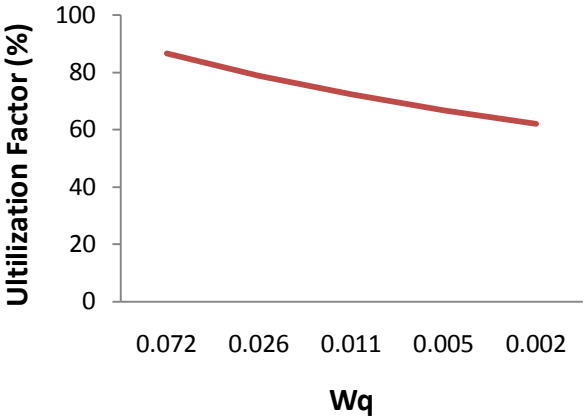

Fig.3: Utilization Factor (ρ) against Average waiting time of patients in the queue(Wq)

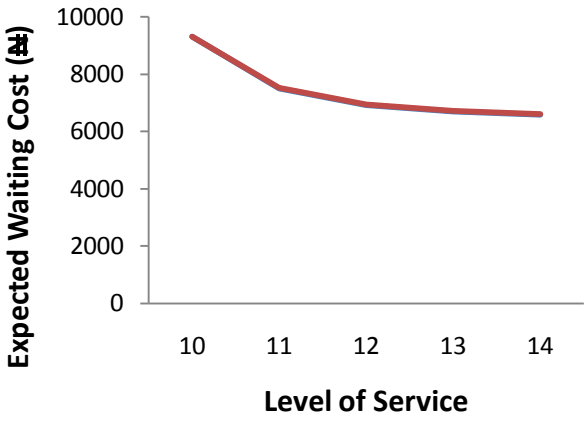

Fig..5: Expected Waiting Cost against Level of Service.

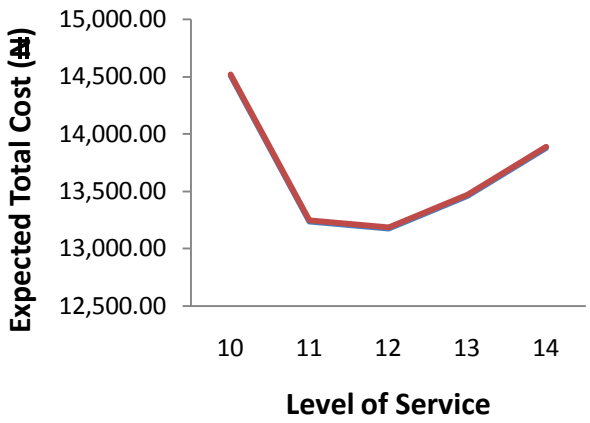

Fig. 6: Expected Total Cost against Level of Service.

4. Discussion of Result

The graphs show that optimal server level at the Clinic is achieved when the number of servers (doctors) is 12 with a minimum total cost of ₦ 13,174.84 per hr as against the present server level of 10 doctors at the Clinic which have high total cost of ₦ 14,509.08 per hr. It should also be noted that patients' average wait time and congestion in the system is also less at this optimal server level.

5. Conclusion

The queuing characteristics at the Riverside Specialist Clinic of the Federal Medical Centre , Makurdi was analysed using a Multi-server queuing Model and the Waiting and service Costs determined with a view to determining the optimal service level. The results of the analysis showed that average queue length, waiting time of patients as well as overutilization of doctors could be reduced when the service capacity level of doctors at the Clinic is increased from ten to twelve at a minimum total costs which include waiting and service costs. The operation managers can recognize the trade-off that must take place between the cost of providing good service and the cost of customers waiting time. Service cost increases as a firm attempts to raise its level of service. As service improves, the cost of time spent waiting on the line decreases. This could be done by expanding the service facilities or using models that consider cost optimization.

REFERENCES

[1] Amakon, U.S ''.How Large is the Opportunity of Queuing in Service Centres? Evidence from Eastern Nigeria''. Department of Economics,Nnamdi Azikiwe University, Awka.2008.

[2] Bastani, P '' A Queuing Model of Hospital Congestion'' .An Unpublished Msc Theses, Department of Mathematics, Simon Fraser University Burnaby,B.C Canada. 2009.

[3] Elegalam "Customer Retention Versus Cost Reduction technique" A Paper Presented at the Bankers Forum held at Lagos, pg.9-10.1978.

[4] Gupta, I., Zoreda, J. and Kramer, N. ''Hospital Manpower Planning by Use of Queuing Theory''. Health Services Research, 6, 76-82.1971

[5] Kandemir-Caues, C., Cauas, L. '', An Application of Queuing Theory to the Relationship between Insulin Level and Number of Insulin Receptors''. Turkish Journal of Biochemistry, 32 (1): 32-38, 2007.

[6] Khan, M.R. and Callahan, B.B. ''Planning Laboratory Staffing with a Queuing Model''. European Journal of Operational Research, 67, 1993.

[7] Kostas, U.N. ''Introduction to Theory of Statistics''. Mc-Gram Hill, Tokyo. 1983.

[8] McClain, J.O. ''.Bed Planning Using Queuing Theory Models of Hospital Occupancy: a Sensitivity Analysis''. Inquiry, 13,167-176,http://www.ncbi.nlm.nih.gov/ 1976.

[9] Ndukwe,H.C,Omale,S and Opanuga O.O ' Reducing Queues in Nigerian Hospital Pharmacy''.African Journal of pharmacy and pharmacology Vol.5(8).pp.1020-1026. 2011.

[10] Nosek, R.A. and Wilson, J.P.'' Queuing Theory and Customer Satisfaction: A Review of Terminology, Trends and Applications to Pharmacy Practice. Hospital Pharmacy'', 36,275-276, http://www.drum.lib.umd.edu. 2001.

[11] Obamiro, J.K. 'Queuing theory and Patient Satisfaction: An overview of terminology & application in Ante-Natal care unit.http://www.upg-bulletin-se.ro

[12] Olaniyi,T.A.'' An Appraisal of Cost of Queuing in Nigerian Banking Sector: A Case Study of First Bank of Nigeria Plc, Ilorin''.Journal of Business & Social Sciences. Vol.9, Nos,1&2, pages 139-145. 2004.

[13] Rosenquist, C.J.'' Queuing analysis: A Useful planning and Management techniques for radiology''. Journal of Medical Systems, 11,413-4, 1987.

[14] Shimshak, D.G., Gropp, D.D. and Burden, H.D. '' A Priority Queuing Model of a Hospital Pharmacy Unit.'' European journal of operational Research .7, 350-354. 1981.

[15] Sharma, J.K. ''A text book on Operations Research; Theory Applications''. 4th Edition. Macmillan publishers, India. 2009.

[16] Stakutis C, Boyle T '' Your Health, your Way: Human-enabled Health Care.'' CA Emerging Technologies, pp. 1-10.2009.

[17] Vikas, S.. '' Use of Queuing Models in Healthcare: Department of Health Policy and Management'', University of Arkansas for Medical science Available at: http://works.bepress.com/vikas-singh/13.2006.

[18] Young, J.P. ''The Basic Model, in a Queuing Theory Approach the Control of Hospital Inpatient Census'', John Hopkins University, Baltimore, 74-79.http/online library. Wiley.com.1962a.

[19] Young, J.P. .''Estimating bed Requirements in a Queuing Theory Approach to the Control of Hospital Inpatient Census,'' John Hopkins University, Baltimore, 98-108. http://online library.wiley.com.1962b.