

## Utility Analysis of an Intensive Care Unit Model Using Queuing Theory with Improved Taguchi Loss Function

Kiprono, D., K.<sup>1</sup>, Rotich, T.<sup>2</sup> and Kerongo, M., J.<sup>1</sup>

1. Department of Mathematics and Actuarial Sciences, Kisii University, P.O Box 408- 40200, Kisii, Kenya.

2. Department of Centre for Teacher Education, (Science Education), Moi University, P.O Box 3900- 30100, Eldoret, Kenya.

**Abstract:** Long queues have affected service delivery in Intensive Care Units of many medical facilities. This study is a utility analysis of a queuing problem at Moi Teaching and Referral Hospital (MTRH) Intensive Care Unit (ICU) in Kenya. The objectives were to determine the average time of a patient in the system, optimum number of beds required and establish the stability of the system. Admission data of ICU for six months was obtained from MTRH. A multi-server queuing Model (M/M/s) was used together with Improved Taguchi Loss Function to analyze the problem and an excel calculator was used to simulate the model results in five scenarios. It was found that the optimum number of beds required in the ICU was 13, which reduces the patient waiting time by 86.06% while server utilization remains good at 77%. Lastly, the stability of the system was found out to be achieved when the bed allocation is between 12 and 14. Therefore, it is recommended that MTRH management, policy makers at county and national level and other health facilities with similar queuing problem improve the overall patient care by installing the optimum number of beds in order to meet the patient needs.

### I. Introduction

Waiting to receive service in a queue occur every day, it affects people in polling stations as they queue to vote, traffic on the road, patients in hospital, customers in shops, buying fuel from a petrol station, queuing on the bank Automatic Teller Machine (ATM), or making withdrawals or deposits in a bank that still require customers to queue physically. Though, currently we must appreciate the automated customer queuing in most banks in Kenya where customers get their number in relation to the type of service required and simply wait for their turn to be announced. This changes the bank scenario to a Multi-Phase, Multi Server queuing system.

[1] studied the waiting line in Ante natal care for expectant women which resulted in evaluating the effectiveness of a queuing model in identifying challenges of expectant mothers in the facilities though, the greatest challenge was the waiting time of each mother to receive service depending on her arrival. A study done by [2] examined a queuing model effectiveness in identifying provider staffing patterns to reduce patients who leave without being seen and their findings was that queuing models are the most effective in determining staff allocation. As a result of reorganizing server capacity over time, customer waiting time reduced and complaints immediately reduced without an addition of staff [2].

[2] state that most systems currently are subject to time-varying demand, where arrival rates and the number of servers vary throughout the period of operation. [3] Studied how an increase in patient arrival rate determined the amount of waiting times and queue length for an emergency radiology service. A system with congestion discourages arrivals in any service facility. [3] Further suggests scheduling patients when possible and segregating patients based on expected examination duration. Such measures would lessen inconsistency and decrease expected waiting times. Hence, waiting time is a determinant of customer satisfaction as analysed by [4] where he concluded that the cost of a dissatisfied customer is not negligible and described Waiting in line as a primary source of dissatisfaction. They also mentioned that queuing theory with Taguchi Loss Function, is useful in deriving costs associated with customer dissatisfaction and that this dissatisfaction is not just an issue at the lower specification limit and the upper specification limit, but rather for each moment in time beyond the targeted wait time. [5] Set out to determine the capacity with minimal costs required to serve patients at the Duke University medical centre. They found that the current capacity is good but needs to be redistributed in time to accommodate patient arrival patterns. A high level of service will cost more to provide services and the service provider may not be able to break even. The amount of work in the system does not depend on the order in which the customers are served [6]. The amount of work decreases with one unit per unit of time independent of the customer being served and when a new customer arrives, the amount of work is increased by the service time of the new customer. Two major costs are therefore necessary to make decision.

## II. Indentations and Equations

The following assumptions were made for the queuing system at MTRH which is in accordance with the queuing theory. They are;

Arrivals follow a Poisson probability distribution at an average rate of  $\lambda$  customers (patients) per unit of time.

The queue discipline is First-Come, First-Served (FCFS) basis by any of the servers and there is no balking or reneging. There is minimal priority classification for some extremely critical arrivals but not significantly affecting the services.

Service times are distributed exponentially, with an average of  $\mu$  patients per unit of time.

There is no limit to the number of the queue (infinite).

The service providers are working at their full capacity.

The average arrival rate is greater than average service rate. This is necessary to create a queue.

Servers here represent doctors, beds, theatre, ICU equipment and other medical personnel necessary to provide full services to the ICU patients.

Service rate is independent of line length; service providers do not go faster because the line is longer.

A model satisfying the above assumptions has the capacity to capture all the parameters that involve a multi-channel server system, where clients are served in a parallel server system. The waiting customers in a queue can be fully served if they are attended by any one of the available servers. With these conditions, the most appropriate model adopted for this work is the Multi-server Queuing model (M/M/s) that can capture the dynamics of an emergency medical service with respect to utility of ICU resources.

Following the characteristics of the hospital emergency service, and the assumptions of the model, the following flow chart represents the flow of patients in the queuing system. The system is illustrated to include  $s$  servers, one queue and a general ward facility for recuperating patients. In this study, the patients either admitted directly to the general ward are not considered to be in the queue, and those discharged from ICU are assumed to have left the system. Also, in case a patient admitted in the general ward becomes seriously sick and require ICU services, it is assumed that the patient will join the queue for the services. Patients in the queue are not necessarily waiting in the bench, but could be admitted in the general ward as they wait for space in the ICU facility.

### Model equations

In this model, the probability of having no patient in the system is given by;

$$P_0 = \left[ \sum_{n=0}^{s-1} \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \left( \frac{s\mu}{s\mu - \lambda} \right) \right]^{-1} \quad (1)$$

And the expected (average) number of customers in the system denoted by  $L_s$  will be,

$$L_s = \frac{\lambda \cdot \mu \left( \frac{\lambda}{\mu} \right)^s}{(s-1)!(s\mu - \lambda)} P_0 + \frac{\lambda}{\mu} \quad (2)$$

While the expected (average) number of customers waiting in the queue  $L_q$  is,

$$L_q = \frac{\lambda \cdot \mu \left( \frac{\lambda}{\mu} \right)^s}{(s-1)!(s\mu - \lambda)} * P_0 \quad (3)$$

In order to check the survival of patients, the necessary parameter, is the average time a customer spends in the system defined as,

$$W_s = \frac{L_s}{\lambda} = \frac{\mu \left( \frac{\lambda}{\mu} \right)^s}{(s-1)!(s\mu - \lambda)} * P_0 + \frac{1}{\mu} \quad (4)$$

Before a patient is served, the patient is expected to wait in the queue defined as,

$$W_q = \frac{L_q}{\lambda} = \frac{\mu \left( \frac{\lambda}{\mu} \right)^s}{(s-1)!(s\mu - \lambda)} \cdot P_0 \quad (5)$$

with the chances of having to wait given by the proportion defined in form of a probability as;

$$p(n \geq s) = \frac{\mu \left( \frac{\lambda}{\mu} \right)^s}{(s-1)!(s\mu - \lambda)} \cdot P_0 \quad (6)$$

The utilization factor ( $\rho$ ). The fraction of time when beds are occupied

$$\rho = \frac{\lambda}{\mu s} \quad (7)$$

The utilization rate of the servers is defined by  $\rho = \frac{\lambda}{s\mu}$  and thus the efficiency of M/M/s model is obtained from the ratio,

$$= \frac{\text{Average number of customers served}}{\text{total number of customers}}$$

In order to evaluate and determine the optimum number of servers in the system, two costs must be considered in making these decisions:

Denote the expected service cost by,

$$E(SC) = sC_s \quad (8)$$

where  $s$  is the number of servers and  $C_s$  is the service cost for each server, let the expected waiting cost of the system be

$$E(WC) = \lambda W_s C_w \quad (9)$$

where;  $\lambda$  is the arrival rate,  $W_s$  is the average time an arrival spends in the system and  $C_w$  is the opportunity cost of waiting by customers.

Adding equation (3.23) and (3.24) yields

$$\begin{aligned} E(TC) &= E(SC) + E(WC) \\ TC &= sC_s + \lambda W_s C_w \end{aligned} \quad (10)$$

Then the results of this model were run using the excel calculator software.

The expected total cost of the queuing model with (1, 2, 3, ... s) servers will be calculated and tabulated. Later, the results will be plotted on a graph to get the equilibrium point of optimum service versus costs.

### III. Figures and Tables

The following data was obtained from MTRH showing the bed occupancy, or number of servers and the service and arrival rates of the patients to the ICU.

Number of ICU beds	$n$	=6
Arrival rate of patients	$\lambda$	=5
Service rate per server	$\mu$	=0.5
Average waiting cost	$C_w$	= Ksh 450
Average service cost	$C_s$	= Ksh 400
Total system cost	$TC$	= Ksh 850

Calculating the probability of no patients in the system using (1)

Taking  $s=11$ ,  $\lambda = 5$  and  $\mu = 0.5$

$$p_0 = \left[ \sum_{n=0}^{10} \frac{(5/0.5)^n}{n!} + \frac{(5/0.5)^{11}}{11!} \left( \frac{11 \times 0.5}{11 \times 0.5 - 5} \right) \right]^{-1}$$

$$p_0 = 0.0000247$$

$$\begin{aligned} L_q &= \left[ \frac{1}{(s-1)!} \left( \frac{\lambda}{\mu} \right) \frac{\mu \lambda}{(\mu s - \lambda)^2} \right] p_0 \\ &= \left[ \frac{1}{(11-1)!} \left( \frac{5}{0.5} \right) \frac{0.5 \times 5}{(0.5 \times 11 - 5)^2} \right] 0.0000247 \\ &= 6.821 \end{aligned}$$

Calculating expected (average) number of customers in the system denoted by  $L_s$  using (2)

$$\begin{aligned} L_s &= \frac{\lambda \mu \left( \frac{\lambda}{\mu} \right)^s}{(s-1)!(s\mu - \lambda)^2} p_0 + \frac{\lambda}{\mu} \text{ or } L_s = L_q + \frac{\lambda}{\mu} \\ &= \frac{5 \times 0.5 \left( \frac{5}{0.5} \right)^{11}}{(11-1)!(11 \times 0.5 - 5)^2} 0.0000247 + \frac{5}{0.5} \\ &= 16.82 \end{aligned}$$

Calculating average time a customer spends in the system using (4)

$$W_s = \frac{L_s}{\lambda}, \quad = \frac{16.82}{5} = 3.364$$

Calculating the average expected patient wait time in the queue using (5)

$$W_q = \frac{L_q}{\lambda}, \quad = \frac{6.821}{5} = 1.3642$$

Calculating the chances of having to wait given by the proportion defined in form of a probability as;

$$p(n \geq s) = \frac{\mu \left(\frac{\lambda}{\mu}\right)^s}{(s-1)!(s\mu - \lambda)} \cdot p_0$$

$$= \frac{0.5 \left(\frac{5}{0.5}\right)^{11}}{(11-1)!(11 \times 0.5 - 5)} * 0.0000247 = 0.681$$

Utilization factor ( $\rho$ ), representing the time the beds are occupied using (7)

$$\rho = \frac{\lambda}{\mu s}, \quad = \frac{5}{0.5 \times 11} = 0.9090909$$

Developing excel calculator using the above model equation and using to run the data.

Calculating performance of 11 beds using the excel calculator was as follows;

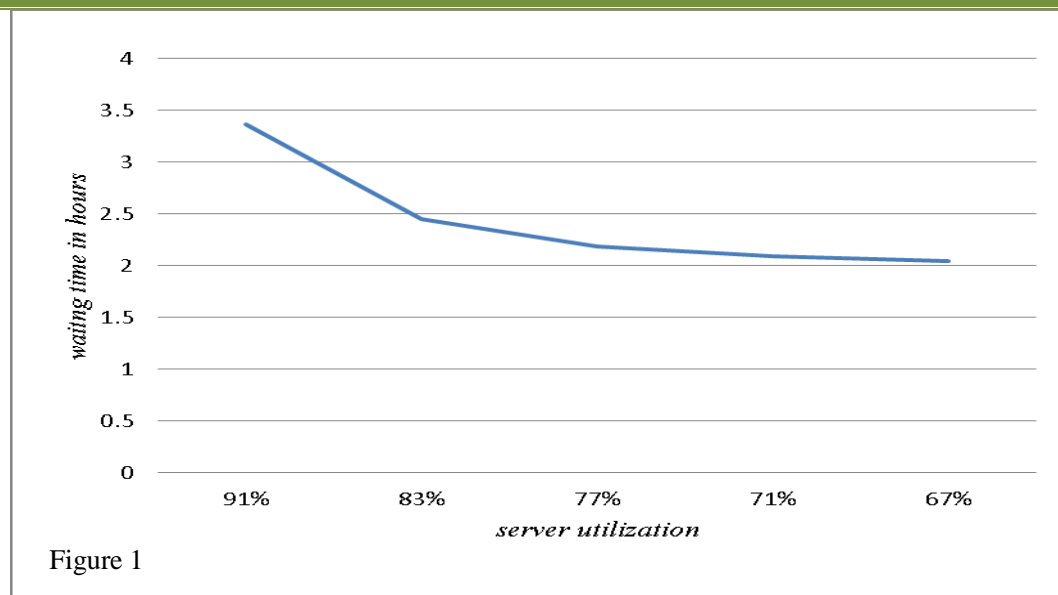
Table 1 Performance of 11 Beds using the Excel Calculator

Parameter	Value	Unit
Arrival Rate (lambda)	5	customers/hour
Service Rate per Server ( $\mu$ )	0.5	customers/hour
Number of Servers	11	servers
Average time between arrivals	0.2	hour
average service time per server	2	hour
combined service rate ( $s \times \mu$ )	5.5	customers/hour
Rho (average server utilization)	0.9090909	
Po (9) Probability the system is empty)	0.00002	
L (average number in the system	16.821182	customers
Lq (average number waiting in the queue)	6.821182	customers
W (average time in the system)	3.3642364	hour
Wq (average time in the queue)	1.3642364	hour

Table 2 Performance Measures of the Model in Five Scenarios

No. of beds	$\lambda$	$\mu$	$p_0$	$\rho$	$L_s$	$L_q$	$W_s$	$W_q$	$P_w$
11	5	0.5	0.000025	90.9	16.82	6.821	3.364	1.364	0.682
12	5	0.5	0.000036	83.3	12.247	2.247	2.249	0.449	0.449
13	5	0.5	0.000041	76.9	10.951	0.951	2.190	0.190	0.285
14	5	0.5	0.000043	71.4	10.435	0.435	2.087	0.087	0.174
15	5	0.5	0.000044	66.7	10.204	0.024	2.041	0.041	0.102

Comparing Waiting Time Against Server Utilization



Determining the Equilibrium Point and the Optimum Number of Beds

Calculating System Costs

Cost of waiting for service per our = ksh 450

Cost of offering the service per hour = ksh 400

Calculating the expected service cost using (8)

$$E(SC) = sC_s$$

$$C_s = ksh\ 400$$

$$E(SC) = sC_s = 11 \times 400 = 4400$$

Calculating waiting cost of the system using (9)

$$E(WC) = \lambda W_s C_w$$

$$= 5 \times 3.364 \times 450 = 7569.53$$

Calculating the Expected Total Costs of the system using (10)

$$E(TC) = E(SC) + E(WC)$$

$$E(TC) = SCS + (\lambda W_s) C_w$$

$$= 4400 + 7569.53 = 11969.53$$

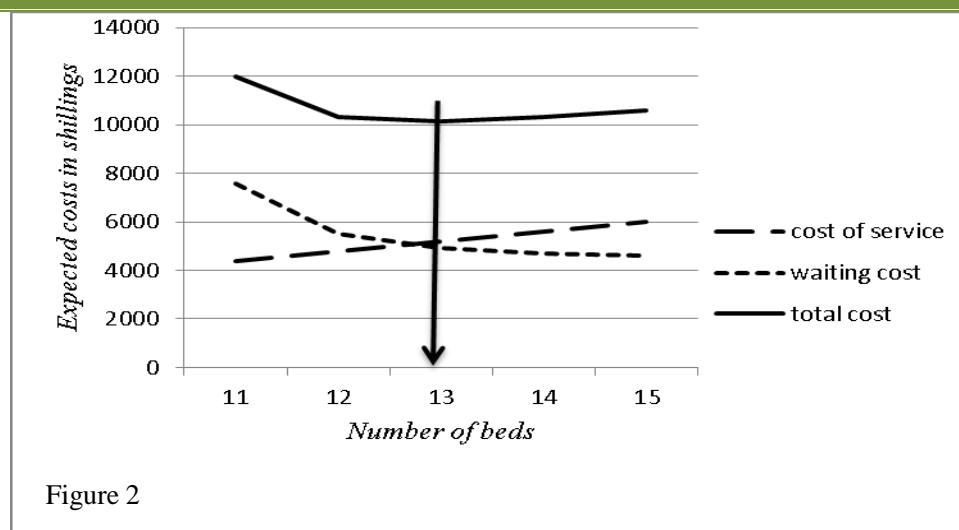
An excel calculator was again developed to compute the expected cost

Computed costs were as follows;

Table 3 system costs in the System against Number of Beds

No.of beds	$\lambda$	$\mu$	$W_s$	$C_s$	$C_w$	$E(SC)$	$E(WC)$	$E(TC)$
11	5	0.5	3.36424	400	450	4400	7569.53	11969.53
12	5	0.5	2.44939	400	450	4800	5511.12	10311.12
13	5	0.5	2.19018	400	450	5200	4927.91	10127.91
14	5	0.5	2.08707	400	450	5600	4695.90	10295.90
15	5	0.5	2.04082	400	450	6000	4591.84	10591.84

Optimum Number of Beds Required using System Costs



#### Determining the Stability of the System

Assuming that each side is willing to tolerate loss of up to Ksh 300, in the total costs and considering values of total costs of the five scenarios calculated, we generate values that will be fitted into the Improved Taguchi Loss graph.

Assuming that a customer is willing to wait for a maximum of 30 minutes without complaining, and the facility does not wish to have an idle bed but wish to have 100% utilization without over straining.

Table 4 Expected Total Costs

Beds	$\lambda$	$\rho$	$\mu$	SC	WC	E(SC)	E(WC)	E(TC)
11	5	90.9	0.5	400	450	4400	7569.53	11969.53
12	5	83.3	0.5	400	450	4800	5511.12	10311.12
13	5	76.9	0.5	400	450	5200	4927.91	10127.91
14	5	71.4	0.5	400	450	5600	4695.90	10295.90
15	5	66.7	0.5	400	450	6000	4591.84	10591.84

Table 5 Cost of Rejection

No of beds	Total Expected costs	Cost of Rejection
11	11969.53	1841.62
12	10311.12	183.21
13	10127.91	0.00
14	10295.90	167.99
15	10591.84	463.93

Tolerance level of customers using Improved Taguchi Loss Function

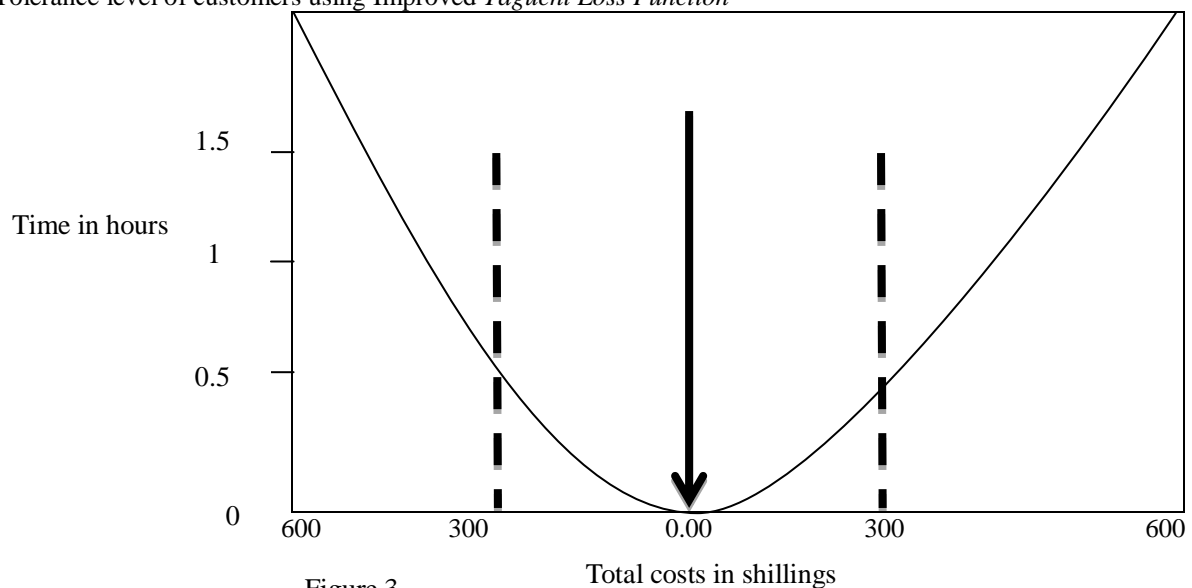


Figure 3

#### IV. Conclusion

The study analysed data from Moi, Teaching and Referral Hospital with an objective of addressing the queuing problem in the Intensive Care Unit that has six ICU beds. It was found that the optimum number of beds required in the ICU was 13. This was achieved when the total system cost was Ksh 10127.91 being the minimum in all the scenarios. Findings also show that, with 13 beds, the patient waiting time reduces by 86.06% while server utilization remains good at 77%. Lastly, the stability of the system was found out to be achieved when the bed allocation is between 12 and 14 by using the total expected costs together with improved Taguchi Loss Function. Therefore, from the findings of this work, it is recommended that MTRH management, policy makers at county and national level and other health facilities with similar queuing problem improve the overall patient care by installing the optimum number of beds in order to meet the patient needs to reduce mortality rates resulting from emergency service provision.

#### Acknowledgement

My gratitude goes to my dear wife, and my children for their prayers and financial support they gave unto me. Also to my supervisors, Dr. Kerongo Joash and Dr. Rotich Titus for their continuous guidance, thank you. Lastly I thank the management of Moi Teaching and Referral Hospital and Kisii University for they are part of this work.

#### References

- [1]. Obamiro, K. (2010). Queuing theory and Patient Satisfaction: An overview of terminology & application in Ante-Natal care unit. [http://www.upg-bulletin se.ro](http://www.upg-bulletin.se.ro)
- [2]. Green L, Kolesar P.J., and Whitt W (2007). Coping with time-varying demand when setting requirements for service systems. *Production and Operations Management*.
- [3]. Rosenquist, C. (1987). Queuing analysis: a useful planning and management technique for radiology. *Journal of Medical Systems*.
- [4]. Gillett, F. (2006). Queuing Theory and the Taguchi Loss Function: The Cost of Customer Dissatisfaction in Waiting Lines. *International Journal of Strategic Cost Management*.
- [5]. Keller, F. & Laughunn, J. (1993). An application of queuing theory to a congestion problem in an outpatient clinic. *Decision Sciences. Matematik*.
- [6]. Resing, A., Ivan, A., & Jacques, (2015). *Queueing Systems*. Eindhoven, The Netherlands: Eindhoven University of Technology.