African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 35-53)
www.abjournals.org

OPTIMIZATION OF A NETWORK OF QUEUES IN A UNIVERSITY TEACHING HOSPITAL

Nwankwo Chike H[1]. and George Ekemini U[2].

[1]Department of Statistics, Nnamdi Azikiwe University, Awka, Nigeria
[2]Department of Statistics, Akwa-Ibom State University, Ikot Akpaden, Nigeria

ABSTRACT: *This paper is aimed at studying the waiting and service times of patients at the Medical Out-Patients Department (MOPD) of the University of Uyo Teaching Hospital, Akwa-Ibom State, Nigeria. The MOPD of the hospital consists of three Units – the Card Room Section, the Vital Signs Unit and the Consulting Rooms. There is a single-server-single-queue system at the Card Room; Two-server-single-queue at the Vital Signs Unit and a seven-parallel single-server queues at the Consulting Rooms. These, put together, form a network of queues. Apart from the inter-arrival times at the Card Room, the inter-arrival times at other sections as well as the service times in all followed distributions other than exponential, and these resulted in models that made use of approximation techniques. Different models were hypothesized at the Vital Signs Unit and the Consulting Rooms, and were combined to see which combination will be most effective at reducing the waiting times of patients in the MOPD. It was realized that a combination of a two-server-single-queue at the Vital Signs Unit and a seven-server-single-queue at the Consulting Room resulted in the least mean waiting times of patients in the MOPD, with the mean waiting time in the queue and in the system being 22 minutes, 15 seconds and 38 minutes, 18 seconds respectively.*

KEYWORDS: Queue, Inter-Arrival, Service, Lognormal, Loglogistic, Coefficient of Variation

INTRODUCTION

A queue is, simply, a line for service. Queues are found in petrol stations, where the automobiles are the customers and the pumps are the service facilities; airports, where the aircrafts are the customers and the runways are the service facilities; post offices, where the letters are the customers and the sorting systems are the service facilities. It is, generally, formed at any place when a customer (human beings or physical entity) that requires service is made to wait due to the fact that the number of customers exceeds the number of service facilities or when service facilities do not work efficiently and take more time than prescribed to serve a customer. Clearly, there are service systems for which we never expect to wait. For instance, a man whose house is on fire wouldn't like to be put on hold (join a queue), while trying to get an emergency response unit. Therefore, queues are seen as a social phenomenon and of serious concern; hence, the need for it to be studied.

Though queues, as a concept, is as old as man's existence, the first time the concept was brought to consciousness was when Erlang (1909) published his work "The Theory of Probability and Telephone Conversation" (Bhat; 2008). After this application, many were exposed to the efficacy of queueing theory in solving problems, especially, in the telephony.

In recent times, the application of queueing theory has been extended to all aspects of human endeavor where orderliness is desired. These include hospitals.

Just as in other scenarios where customers seek to be served, long queues are undesirable in the hospitals because delay in receiving needed services can cause prolonged discomfort and economic loss (when patients are unable to work); and possible worsening of their medical conditions which can increase subsequent treatment costs and poor health outcomes. In extreme cases, long queues can delay diagnoses and (or) treatments, to the extent that death may occur while a patient waits.

As a result of these long queues, patients, most times, fail to access these services due to the fact that they are, most times, being subjected to wait for a long time to obtain the service. They consider waiting in a queue for a long time as a waste of time, especially when they are faced with the need to attend to some other ‘seemingly’ important issues. There is, therefore, a need to improve on the accessibility of these facilities by way of reducing the waiting time of those who require the services, as well as balance cost of rendering medical service so as to achieve a win-win situation for both service provider and service user.

In recent times, queueing theory has been applied in many areas to help improve services rendered in facilities. It has proven to be an efficient tool for managing the waiting times of customers in service facilities. For applications in Supermarkets, see Jhala and Bhathawala, 2017: Igwe, et al, 2014. In the Banking sector, Abdul-Wahab and Najim (2014) applied Queueing Theory to the Automated Teller Machine (ATM) for service optimization at a bank branch in Ghana. Jhala and Bhathawala (2016), using the M/M/c model, determined the optimum number of servers for a bank branch in India. In the Health Sector, Kembe et al (2012) studied the queueing characteristics at the Riverside Specialist Clinic of the Federal Medical Center, Makurdi using the multi-server queueing model; Adeleke et al (2009) considered the waiting time of patients in the University of Ado-Ekiti health center as a single-channel queueing system.

This study seeks to compare the performance measures at the three sections of the Medical Out-Patient Department of the University of Uyo Teaching Hospital (Card Room, Vital Signs Unit and the Consulting Rooms sections) in Nigeria. Here the Card Room operates as a single-queue-single-server system; the Vital Signs Unit operates as a single-queue-multi(two)-server system, while the Consulting Rooms operate as seven parallel single-server-single-server systems. Effort will be made to input the true models of arrival and service times for more accurate results. An optimum number of service channels at the Consulting Rooms will also be obtained.

METHODS

Data Collection

Data for this work were collected by tracking randomly chosen patients from the Card Room through the Vital Signs Unit and the Consulting Rooms. Patients were tracked for all the seven Consultants. In all, 250 patients were tracked.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 35-53)
www.abjournals.org

Computing the Performance Measures

The decomposition method was employed in analyzing the network of queues. The different nodes in the network are decomposed into subnetworks and for each of the nodes, the mean arrival rate and the mean service rate are determined (the inter-arrival and service times distributions, as well); then, the node is analyzed as a single queue to obtain the performance measures. When one of the performance measures is obtained, the Little's formula (Little,1961) can be used to obtain others.

The mean queue waiting time for a single server system W_q is computed using the formula;

$W_q = \frac{\rho}{\mu-\lambda}\frac{C_A^2+C_S^2}{2}$ (Gross, et al. 2008). . . 1

whereρ is the traffic intensity, λ is the mean arrival rate, μ is the mean service rate, C_A^2 and C_S^2 are, respectively, the squared coefficients of variation for the inter-arrival and service times.

Equation 1 reduces to

$W_q = \frac{\rho}{\mu-\lambda}$ . . . 2

Since C_A^2 for a poisson process and C_S^2 for exponentially distributed variables are unities

Equation (1), popularly known as Allen-Cunneen approximation formula, is exact for the M/M/1 and M/D/1 model. (Gross, et. al.: 2008).

When the number of servers $c \geq 2$, equation (1) become;

$W_q \approx \frac{r^c}{c!c\mu(1-\rho)^2} p_0 \frac{C_A^2+C_S^2}{2}$ . . 3

Where $r = \frac{\lambda}{\mu}$

andp_0, the steady-state probability of having no patient in the system is

$p_0 = \left\{\sum_{n=0}^{c-1}\frac{r^n}{n!} + \frac{c^c}{c!}\sum_{n=0}^{\infty}\left(\frac{r}{c}\right)^n\right\}^{-1}$ (Gross et al., 2008)

The mean waiting time in the system is computed as

$W \approx \frac{1}{\mu} + W_q$ (Sharma, 2010). . . 4

The expected queue length (L_q) and the expected number of patients in the system (L) can then be computed using the Little's formulas viz

$L_q = \lambda W_q$ . . . 5

and

$L = \lambda W$ . . . 6

Comparing Models

At the Consulting Rooms, there are seven servers with seven parallel queues. A hypothetical model is considered by merging the seven queues to form a single queue with multiple servers. The performance measures of this hypothetical model are evaluated, to see which is more effective at reducing the waiting times of patients, at the Consulting Rooms.

The mean queue waiting time is computed using equation (3) and other performance measures would be obtained as before.

Finding the Optimum Number of Servers

Queues are formed whenever the number of customers (patients in this case) is more than the available servers; or when the arrival rate to a system is more than the service rate. Increasing the number of servers in a system, definitely, reduces the waiting time of customers in the queue, thereby satisfying the customers. This, invariably, means extra service cost on the service providers. Since resources are scarce, there is therefore a need to strike a balance between customers' satisfaction and the cost of service. This can be achieved in a system by finding the number of servers that would bring about this trade-off.

In order to determine the optimum number of servers at the consulting rooms, four scenarios (7 servers, 8 servers, 9 servers, and 10 servers) were considered; and two types of costs – service cost and waiting cost, were also considered.

Let k_1 and k_2 denote the waiting cost and service cost per unit time, respectively; and let C_T denote the total cost, then;

Expected total costs = Expected Waiting cost in the system + Expected Service cost.

$E(C_T) = (\lambda W)k_1 + ck_2;$

which can be written as

$E(C_T) = Lk_1 + ck_2$ (Kembe *et al.*, 2010) . . 7

Data Presentation and Model Development

Below are the frequency tables and graphs of the inter-arrival and service times at the different units of the facility.

At the card room, the total inter-arrival times for the 250 patients was 45420seconds, having a mean of 181.68seconds and a variance of 22916.04seconds.

Table 1 shows measures of the service times at the card room, vital signs unit and consulting rooms.

Table 1: Summary of Service Times at Other Units of the MOPD

Unit of MOPD	No of Patients	Total Service Time(seconds)	Mean Service Time (sec)	Variance
Service at the Card Room				
Service at the Vital Signs Unit	250	39913	159.65	6876.68
Service at the Consulting Rooms	250	71891	287.56	12249.32
	250	244069	976.28	121158.60

Table 2: Frequency Table for the Inter-Arrival time at the Card Room

CLASS (Seconds)	MID-VALUE	FREQUENCY
0 – 50	25	50
50 – 100	75	51
100 – 150	125	27
150 – 200	175	32
200 – 250	225	20
250 – 300	275	21
300 – 350	325	14
350 – 400	375	11
400 – 450	425	7
450 – 500	475	8
500 – 550	525	2
550 – 600	575	3
600 – 650	625	2
650 – 700	675	1
700 – 750	725	1

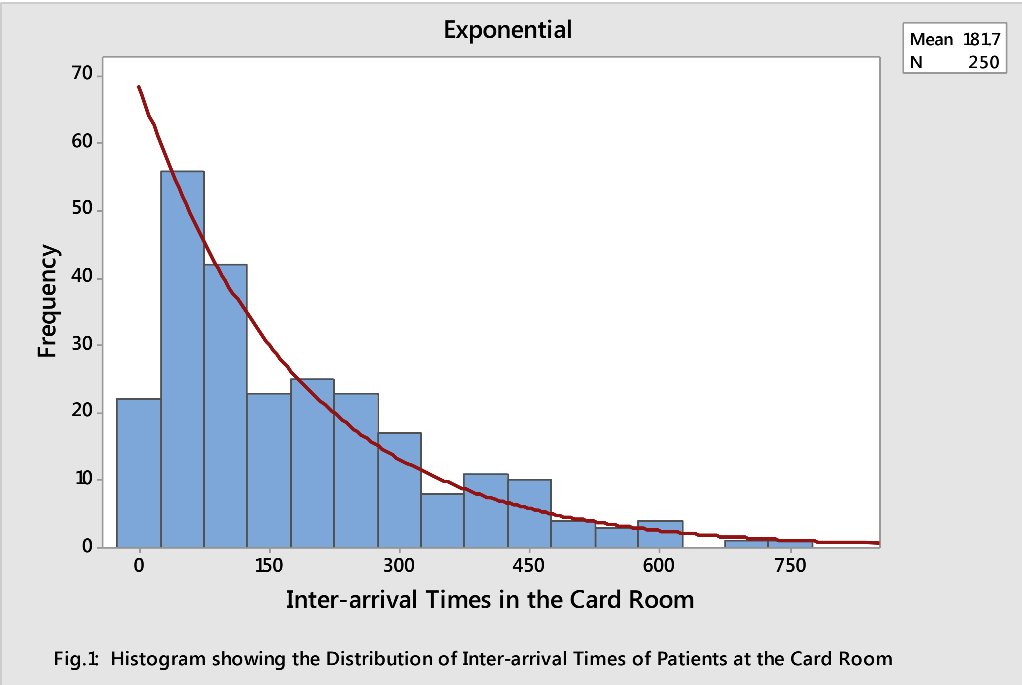

Fig. 1: Shows the Graph of the Inter-Arrival Times at the Card Room.

The graph shows an exponential decay thereby suggesting an exponential distribution. A Chi-Square goodness of fit test carried out on the data had a p-value of 0.148, which is a strong evidence that the distribution of the inter-arrival times at the Card Room in exponential.

Table 3: Frequency Table for the Service Times at the Card Room

CLASS (Seconds)	MID-VALUE	FREQUENCY
25 – 75	50	6
75 – 125	100	94
125 – 175	150	64
175 – 225	200	53
225 – 275	250	14
275 – 325	300	10
325 – 375	350	3
375 – 425	400	0
425 – 475	450	4
475 – 525	500	1
525 – 575	550	0
575 – 625	600	0
625 – 675	650	0
675 – 725	700	0
725 – 775	750	0
775 – 825	800	1

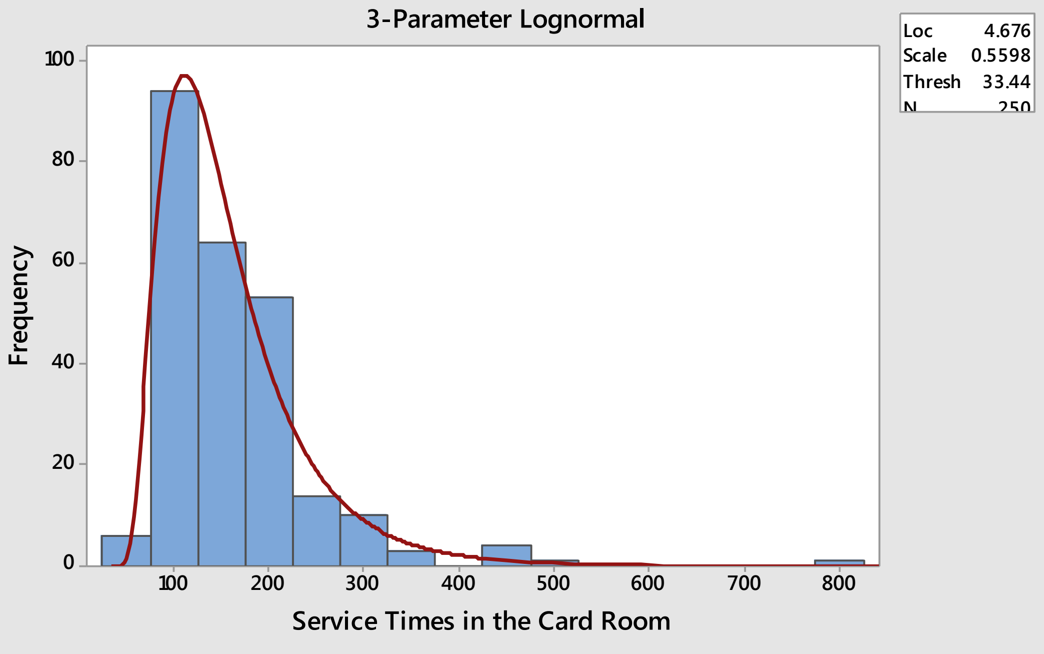

Fig. 2a: Histogram showing the Distribution of Service Times of Patients at the Card Room

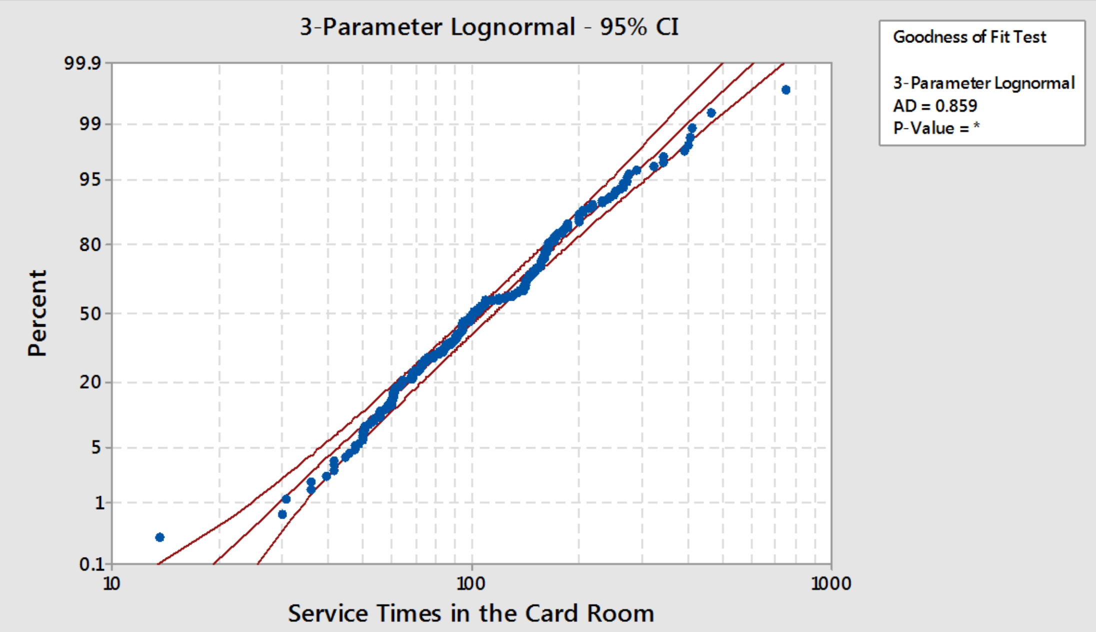

Fig. 2b: Probability Plot for Service Times in the Card Room

Fig. 2: Suggests that the 3-Parameter Lognormal Distribution would fit it well.

A close look at the graph of the service time. The Anderson-Darling goodness of fit test showed a value of 0.859, with a p-value of 0.027. The null hypothesis of a good fit is not rejected at 1% level of significance. The lognormal distribution is thus adopted for the service times at the Card Room.

The probability density function of the 3-parameter lognormal distribution is given as;

$$f(x;\mu,\sigma^2,\gamma) = \frac{1}{(x-\gamma)\sqrt{2\pi\sigma^2}} e^{\frac{-1}{2\sigma^2}(\ln x-\gamma-\mu)^2}, \gamma < x < \infty, \mu > 0, \sigma^2 > 0 \,. \qquad 8$$

The mean and variance functions of the distribution are respectively given as

$$E(X) = \gamma + e^{\mu}\sqrt{\omega} \qquad 9$$

and

$$V(X) = e^{2\mu}\omega(\omega-1) \qquad ;\text{ where } \omega = e^{\sigma^2} \qquad 10$$

The parameters of the distribution are estimated from the data, by solving the set of equations given by (Cohen and Whitten, 1980). These estimated values are then substituted into the mean and variance functions to obtain the mean and variance that go into computing the squared coefficient of variation.

Table 4: Frequency Table for the Service Times at the Vital Signs Unit

CLASS (Seconds)	MID-VALUE	FREQUENCY
125 – 175	150	12
175 – 225	200	50
225 – 275	250	75
275 – 325	300	48
325 – 375	350	38
375 – 425	400	10
425 – 475	450	5
475 – 525	500	3
525 – 575	550	3
575 – 625	600	0
625 – 675	650	2
675 – 725	700	0
725 – 775	750	1
775 – 825	800	1
825 – 875	850	0
875 – 925	900	0
925 – 975	950	2

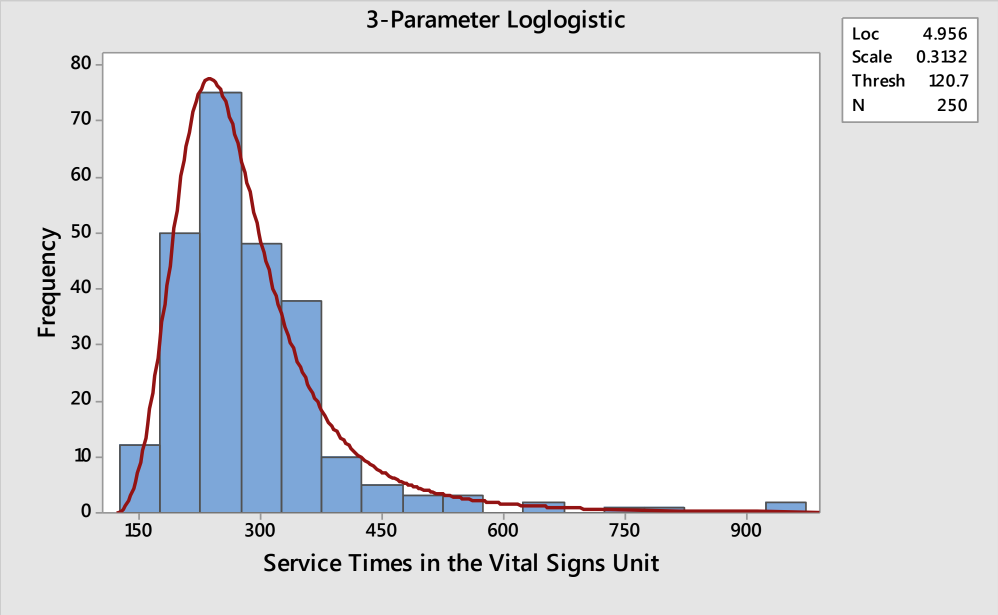

Fig. 3a: Histogram showing the Distribution of Service Times of Patients at the Vital Signs Unit

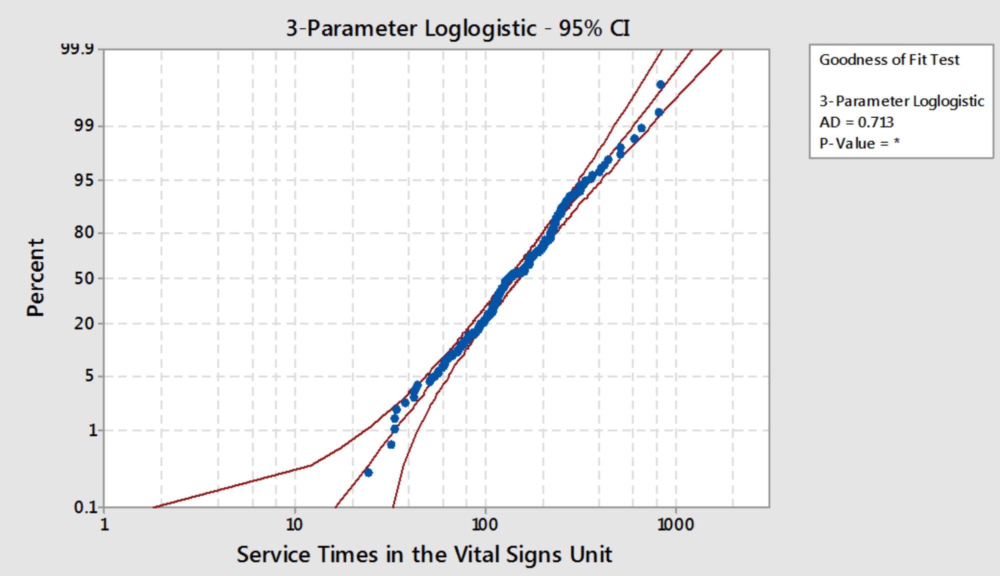

Fig. 3b: Probability Plot for Service Times in the Vital Signs Unit

Fig. 3: Suggests not the usual Exponential Distribution; rather, it suggests the 3-Parameter Loglogistic Distribution.

Again, a close look at the graph of the service times data at the Vital Signs Unit, The Anderson-Darling goodness of fit showed a value of 0.713, with a p-value of 0.063. The null hypothesis of a good fit is not rejected at 1% level of significance. The 3-parameter loglogistic distribution is thus adopted for the service times at the Vital Signs Unit. Its probability density function is given as

$$f(x;\alpha,\beta,\gamma) = \frac{\frac{\beta}{\alpha}\left\{\frac{x-\gamma}{\alpha}\right\}^{\beta-1}}{\left\{1+\left[\frac{x-\gamma}{\alpha}\right]^{\beta}\right\}^{2}}, x > \gamma, \alpha > 0, \beta \geq 1 \qquad 11$$

The mean and variance function are respectively given as

$$E(X) = \gamma + \alpha B(1+\beta^{-1}, 1-\beta^{-1}) \qquad 12$$

and

$$V(X) = \alpha^2[B(1+2\beta^{-1}, 1-2\beta^{-1}) - B^2(1+\beta^{-1}, 1-\beta^{-1})] \qquad 13$$

where $B(.,.)$ is he beta function.

The parameters of the distribution are estimated from the data, by solving the set of equations given by (Singh *et al.*, 1993). These estimated values are then substituted into the mean and variance functions to obtain the mean and variance that go into computing the squared coefficient of variation.

Table 5: Frequency Table for the Service Times at the Consulting Rooms

CLASS (Seconds)	MID-VALUE	FREQUENCY
150 – 250	200	1
250 – 350	300	1
350 – 450	400	4
450 – 550	500	8
550 – 650	600	22
650 – 750	700	18
750 – 850	800	29
850 – 950	900	76
950 – 1050	1000	41
1050 – 1150	1100	7
1150 – 1250	1200	2
1250 – 1350	1300	3
1350 – 1450	1400	2
1450 – 1550	1500	9
1550 – 1650	1600	3
1650 – 1750	1700	9
1750 – 1850	1800	11
1850 – 1950	1900	2
1950 – 2050	2000	1
2050 – 2150	2100	1

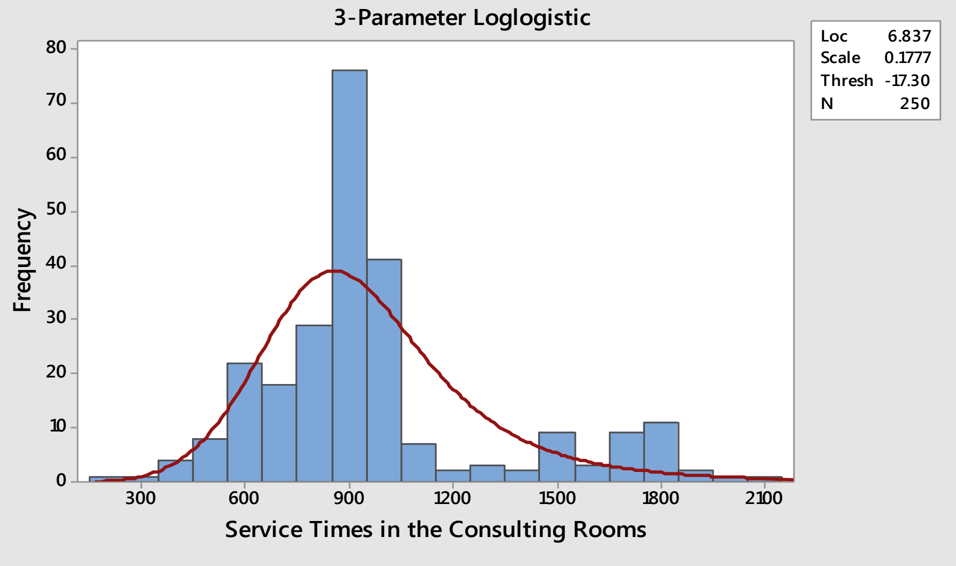

Fig. 4: Histogram showing the Distribution of Service Times of Patience at the Consulting Rooms

Fig 4: Clearly shows that the Distribution of the Service Times at the Consulting Rooms is not Exponential; and is not close to any known Distribution.

It is obvious that the distributions of the inter-arrival and service times are not exponentially distributed (apart from the inter-arrival times at the Card Room). This calls for model approximation. The method used here is the two-moment approximation. To obtain these moments, the parameters of the distributions of the inter-arrival times and service times will be estimated, and used in computing the mean and variance of the distributions, which will go into computing the coefficient of variations.

ANALYSES AND RESULTS

Performance Measures at the Card Room

Because the inter-arrival times are exponentially distributed, and the service times are Lognormal distributed, the model at this unit is the M/LND3/1.

From Table.1, the mean inter-arrival time is 181.68 seconds.

So that the mean arrival rate $(\lambda) = 181.68^{-1} = 0.0055$ patients per second.

Similarly, from Table 1 the mean service time is 159.65 seconds.

Hence, the mean service rate $(\mu) = 159.65^{-1} = 0.0063$ patients per second.

Thus, the traffic intensity (ρ) is

$$\rho = \frac{\lambda}{\mu} = \frac{0.0055}{0.0063} = 0.873$$

Computing the mean and variance of the 3-parameter lognormal distribution, using the estimates of the parameters from the data (Table 1), and equations 9 and 10 it is realized that $s_S^2 = 6934.67$ and $\overline{X}_S^2 = 159.65$.

Thus, the squared coefficient of variation for the card room service time data is

$$C_S^2 = \frac{s_S^2}{\overline{X}_S^2} = \frac{6934.67}{159.65^2} = 0.27$$

Hence, the mean queue waiting time is (using equation 1) as

$$\widehat{W}_q = \frac{\rho}{\mu - \lambda}\frac{(1 + C_S^2)}{2} = \frac{0.873}{0.0063 - 0.0055}\left(\frac{1 + 0.27}{2}\right)$$

$= 693.125$ seconds$\approx$ 11 minutes, 33 seconds.

The mean waiting time in the system is computed, using equation (4) as

$\widehat{W} \approx \frac{1}{\mu} + \widehat{W}_q = 159.65 + 693.125 = 852.775$ seconds $\approx$ 14 minutes, 13 seconds.

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 35-53)
www.abjournals.org

The expected queue length is computed, using equation (5) as

$\hat{L}_q = \lambda\hat{W} = 0.0055(693.125) = 3.81 \approx 4$ patients.

The expected number of patients in the system on the average, at any point in time is computed, using equation (6) as
$\hat{L} = \lambda\hat{W} = 0.0055(852.775) = 4.69 \approx 5$ patients.

Table 6: Summary of Performance Measures for the Card Room

Symbol	λ	μ	ρ	$\hat{W}_q$	$\hat{W}$	$\hat{L}_q$	$\hat{L}$
Value	0.005	0.0063	0.873	11minutes 33seconds	14minutes 13seconds	4patients	5patients

The results of the analysis at the Card Room as summarized in Table 6 reveal that the traffic intensity (which is the measure of the level of congestion in the system) is 0.873. A patient, on arrival at this unit, will have to wait in the queue for about 12 minutes on the average before accessing service; and will spend about 3 minutes in service. This makes the total time a patient spends in the system, on the average, to be about 15 minutes. On the average, there are 4 patients in the queue, and a patient in service, thereby bringing to 5, the total number of patients in the system.

Performance Measures at the Vital Signs Unit

The inter-arrival times at this unit is the same as the service times at the Card Room (which is a 3-parameter lognormal distribution) since no patient can join the queue from outside (that is, without passing through the Card Room). From section 2.5, it is noted that the service times at the Vital Signs Unit follow a 3-parameter loglogistic distribution, and there are two servers at this unit, the model is thus LND3/LLD3/2.

The mean inter-arrival time is the same as the mean service time at the Card Room because patients arrive at the Vital Signs Unit immediately after service at the Card Room.

The mean service time is 287.56 seconds.

The mean arrival rate $(\lambda) = 159.65^{-1} = 0.0063$ patients per second.

The mean service rate $(\mu) = 287.56^{-1} = 0.0035$ patients per second.

The traffic intensity (ρ) is computed as

$$\rho = \frac{\lambda}{c\mu} = \frac{0.0063}{2(0.0035)} = 0.90$$

Computing the mean and variance of the 3-parameter loglogistic distribution, using the estimates of the parameters from the data (Table 1), and equations 12 and 13, it is realized that $s_S^2 = 12237.72$ and $\overline{X}_S^2 = 287.56$.

$$C_S^2 = \frac{s_S^2}{\bar{X}_S^2} = \frac{12237.72}{287.56^2} = 0.148$$

The mean queue waiting time is approximated, using equation (3), ie

$$\widehat{W}_q \approx \frac{r^c}{c!\, c\mu(1-\rho)^2} P_0 \frac{C_A^2 + C_S^2}{2}$$

where $c = 2$ and $r = \frac{0.0063}{0.0035} = 1.8$

and $p_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^{\infty} \left(\frac{r}{c} \right)^n \right)^{-1} = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{c}{c!} \frac{r^c}{(c-r)} \right)^{-1}$

$$= \left(\sum_{n=0}^{1} \frac{1.8^n}{n!} + \frac{2}{2!(} \frac{1.8^2}{2-1.8)} \right)^{-1} = \left(1 + 1.8 + \frac{3.24}{0.2} \right)^{-1} = 0.053$$

$$\frac{r^c}{c!\, c\mu(1-\rho)^2} \approx \frac{1.8^2}{2(2 * 0.0035)(0.1^2)} = \frac{3.24}{0.00014} = 23142.86$$

$$\frac{C_A^2 + C_S^2}{2} = \frac{0.27 + 0.148}{2} = 0.209$$

So that;

$\widehat{W}_q \approx 23142.86 * 0.053 * 0.209 = 256.35$ seconds $\approx$ 4 minutes, 16 seconds.

The mean waiting time in the system is computed, using equation (4) as

$\widehat{W} \approx \frac{1}{\mu} + \widehat{W}_q = 287.56 + 256.35 = 543.91$ seconds $\approx$ 9 minutes, 4 seconds.

The expected queue length is computed, using equation (5) as $\hat{L}_q = \lambda\widehat{W} = 0.0063(256.35) = 1.62 \approx 2$ patients.

The expected number of patients in the system is computed, using equation (6) as $\hat{L} = \lambda\widehat{W} = 0.0063(543.91) = 3.43 \approx 4$ patients.

Table 7: Summary of Performance Measures for the Vital Signs Unit

Symbol	λ	μ	ρ	$\widehat{W}_q$	$\widehat{W}$	$\hat{L}_q$	$\hat{L}$
Value	0.0063	0.0035	0.90	4 minutes 15 secs	9 minutes 4 secs	2 patients	4 patients

At the Vital Signs Unit, there are two servers, working simultaneously to render services to a single queue. The level of congestion, as indicated by the traffic intensity is 0.90. A patient's mean queue waiting time is 4 minutes 16 seconds and the total time spent in the system, on the average, is 9 minutes, 4 seconds. The difference between these two (which is about 5 minutes) gives the average service time of the servers. The average number of patients in the queue is 2 while there are 2 other patients in service, making the average number of patients in the system to be 4.

Supposing the two severs at the Vital Signs Unit were allowed to work as 2 parallel single-queue-single-channel, the following statistics are obtained.

Table 8: Summary of Statistics for the two Servers at the Vital Signs Unit

Server	n	$\bar{x}$	s^2	C_s^2
1	118	290.76	13254.64	0.157
2	132	284.70	11427.48	0.141

Using these statistics and the knowledge that the mean arrival rate at each of these two servers will be half of the mean arrival rate when there is a single queue (that is, for each of the servers, $\lambda = \frac{0.0063}{2} = 0.00315$ patients per second), the following performance measures are obtained for the two servers.

Table 9: Summary of Performance Measures for the Servers at the Vital Signs Unit

Server	1	2
ρ	0.916	0.897
W_q	10 min., 43 sec	8 min., 7 sec
W	15 min., 34 sec	12 min., 52 sec
L_q	2 patients	2 patients
L	3 patients	3 patients

Performance Measures at the Consulting Rooms

Here, the inter-arrival times have the same distribution as the service times in the Vital Signs Unit. Fig. 4 also shows that the service times are not exponentially distributed. There are seven independent servers, serving seven parallel queues. Hence, the models here are seven parallel LLD3/G/1

The Seven Parallel LLD3/G/1 Models

The mean arrival rate at the Consulting Rooms is twice the mean service rate at the Vital Signs Unit (since there are two servers there). The patients arriving here will join one of the seven parallel queues at the rate of one-seventh the rate they arrived from the Vital Signs Unit. Hence, the mean arrival rate at each of the seven (7) rooms is two-seventh of the mean service rate at the vital signs unit, so

$$\lambda = \frac{2(0.0035)}{7} = 0.001$$

African Journal of Mathematics and Statistics Studies
ISSN: 2689-5323
Volume 3, Issue 1, 2020 (pp. 35-53)
www.abjournals.org

Table 10 shows the summary of the mean service times, the variances and the squared coefficients of variation obtained from the data collected at the different consulting rooms (servers).

Table 10: Summary of statistics at the different servers in the Consulting Rooms

Server	n	$\bar{x}$	s_S^2	C_S^2
1	38	1084.34	220110.00	0.187
2	35	891.20	91615.40	0.115
3	40	978.88	75248.73	0.079
4	32	941.72	100763.00	0.114
5	38	993.26	144195.60	0.146
6	36	920.11	109751.40	0.130
7	31	1016.58	92138.12	0.089

Equation (1) was used in approximating the mean queue waiting time for each of the servers. Equations 4, 5 and 6 were used to compute the mean waiting time in the system, the expected queue length and the expected number of patients in the system, respectively. The results are presented in Table 11 below.

Table 11: Summary of Performance Measures at the Consulting Rooms

Server	ρ	$\widehat{W}_q$	$\widehat{W}$	$\hat{L}_q$	$\hat{L}$
1	1.084				
2	0.891	16 min., 19 sec.	31 min., 7 sec.	1 patient	2 patients
3	0.979	92 min., 44 sec.	109 min., 4 sec.	6 patients	7 patients
4	0.942	34 min., 20 sec.	50 min., 3 sec.	2 patients	3 patients
5	0.993	242 min., 34 sec.	259 min., 5 sec.	15 patients	16 patients
6	0.920	23 min., 37 sec.	38 min., 55 sec.	1 patient	2 patients
7	1.017				

Obviously, the traffic intensities for servers 1 and 7 are greater than 1.This implies that the queues at these points would continue to grow without bounds. These servers do not have steady-state solutions. Also worthy of attention is server 5, with traffic intensity approaching 1. The effect of this high value is made obvious in the alarming queue and system waiting times. Servers 2 and 6 are impressive in performance. As a matter of fact, they help to cushion the effects of the excessive waiting times in other servers.

The LLD3/G/7 Model at the Consulting Unit

The aim of this paper is to compare performance measures of the current queues' arrangements in the MOPD of University of Uyo Teaching Hospital and other possible queues combinations, especially where there is more than one server (Vital Signs Unit and Consulting Unit), for optimum performance of that section of the medical facility.

The mean service time is 976.28 seconds, with variance, 121158.60 seconds.

The mean service rate $(\mu) = 976.28^{-1} = 0.00102$ patients per second.

The service time coefficient of variation is $\frac{121158.60}{976.28^2} = 0.127.$

The number of servers $(c) = 7.$

The mean queue waiting time is estimated, using Equation 3 as

$$\widehat{W}_q \approx \frac{r^c}{c!c\mu(1-\rho)^2} P_0 \frac{C_A^2+C_S^2}{2};$$

where

$$c = 7, r = \frac{\lambda}{\mu} = \frac{0.007}{0.00102} = 6.86\ and\ \rho = \frac{r}{c} = \frac{6.86}{7} = 0.98$$

Notice that since there are two servers at the Vital Signs Unit, delivering service at the average rate of 0.0035 patients per second, the average arrival rate at the Consulting Rooms is twice the service rate at the Vital Signs Unit. That is 2(0.0035) = 0.007

$$p_0 = \left(\sum_{n=0}^{c-1} \frac{r^n}{n!} + \frac{c^c}{c!} \sum_{n=c}^{\infty} \left(\frac{r}{c} \right)^n \right)^{-1} = \left(\sum_{n=0}^{6} \frac{6.86^n}{n!} + \frac{7}{7!} \frac{6.86^7}{7-6.86} \right)^{-1}$$

$$= (448.82 + 5888.09)^{-1} = 0.000158$$

$$\frac{r^c}{c!\, c\mu(1-\rho)^2} = \frac{6.86^7}{7!\,(7 * 0.00102)(1-0.974)^2} = \frac{714938.71}{0.0144} = 49648521.53$$

$$\frac{C_A^2 + C_S^2}{2} = \frac{0.148 + 0.127}{2} = 0.1375$$

Therefore

$\widehat{W}_q \approx 49648521.53 * 0.000158 * 0.1375 = 1078.61$ seconds = 17 minutes, 59 seconds.

The mean waiting time in the system $(\widehat{W})$ is computed, using equation 4 is

$\widehat{W} \approx \frac{1}{\mu} + \widehat{W}_q = 1078.61 + 675.67 = 1754.28$ seconds ≈ 29 minutes, 14 seconds.

The expected queue length is computed, using equation 5 as $\hat{L}_q = \lambda \widehat{W}_q = 0.007(1078.61) = 7.55 \approx 8$ patients.

The expected number of patients in the system is computed, using equation 6 as $\hat{L} = \lambda \widehat{W} = 0.007(1754.28) = 12.27 \approx 13$ patients.

Table 12 displays the performance measures for the hypothesized model for the MOPD.

Table 12: Summary of Performance Measures for the LLD3/G/7 model

Symbol	λ	μ	ρ	$\widehat{W}_q$	$\widehat{W}$	$\hat{L}_q$	$\hat{L}$
Value	0.007	0.00102	0.98	17minutes 59 seconds	29 minutes 14 secs	8 patients	13 patients

With the model, the traffic intensity is 0.98; the mean queue and service times are 17 minutes, 5 seconds and 2 minutes, 14 seconds, respectively. The expected number of patients in the queue and in the system are respectively 8 and 13.

Comparing these results to those of table 4 reveals a tremendous improvement. There is a remarkable reduction in the waiting times of patients in the queue and in the system. This, of course, makes the LLD3/G/7 model a better model for the Consulting Rooms of the MOPD.

Finding the Optimum Combination of Models for the MOPD

Table 13 shows the mean waiting times of patients for the different possible combinations of models at the Vital Signs Unit and the Consulting Rooms.

Table 13: Waiting Times for the Different Combinations of Models

VITAL SIGNS UNIT	**CONSULTING ROOMS**	**Seven parallel LLD3/G/1**		**LLD3/G/7**	
Two parallel LND3/LLD3/1		$\widehat{W}_q$	91 min., 20 sec.	$\widehat{W}_q$	27 min., 24 sec.
		$\widehat{W}$	111 min., 52 sec.	$\widehat{W}$	43 min., 27 sec.
LND3/LLD3/2		$\widehat{W}_q$	86 min., 11 sec.	$\widehat{W}_q$	22 min., 15 sec.
		$\widehat{W}$	106 min., 43 sec.	$\widehat{W}$	38 min., 18 sec.

A look at the mean queue waiting times and the mean system waiting times as displayed in Table 13 shows that the best combination is the LND3/LLD3/2 at the Vital Signs Unit and LLD3/G/7 at the Consulting Rooms. As this yields the least mean waiting times in the queue and in the system (22 minutes, 15 seconds and 38 minutes, 18 seconds, respectively). Hence, it can be concluded that this is the combination that optimizes the waiting times of patients at the Medical Out-Patient Department of the Hospital.

Finding the Optimum Number of Servers

The table below shows the summary of the performance measures in the four different scenarios;

Table 14: Summary of the Performance Measures, using Different Number of Servers

c	ρ	$\widehat{W}_q$	$\widehat{W}$	$\hat{L}_q$	$\hat{L}$
7	0.98	1078.61 seconds	1754.28 seconds	7.55 patients	12.27 patients
8	0.85	63.90 seconds	1040.00 seconds	0.44 patients	7.24 patients
9	0.76	21.73 seconds	998.01 seconds	0.15 patients	6.95 patients
10	0.68	8.21 seconds	984.49 seconds	0.06 patients	6.85 patients

Considering four different scenarios give the results in Table 14, when there are 8 servers, it is observed that the queue is almost non-existent. The additions of more servers to the system will introduce unnecessary service cost, because some of the servers would be idle most of the times.

For any two arbitrary values of k_1 and k_2 , as defined in Equation 7

$$E(C_T) = Lk_1 + ck_2,$$

it is apparent that the optimum number of servers is 8. To illustrate this, the expected total cost is evaluated for three different cases (when $k_1 = k_2$, when $k_1 < k_2$, and when $k_1 > k_2$). The results are displayed in Fig 4.1 below.

Fig. 5 below is the graph of the expected total cost against different number of servers, for different values of k_1 and k_2. Increasing the number of servers from seven to eight leads to a reduction in the expected total cost. But, moving from eight servers to nine (and beyond) the expected total cost begins to increase. This shows that beyond eight servers, increasing the number of servers only succeeds in introducing unnecessary cost of service in the system, without any significant reduction in the cost of waiting; thereby making the optimum number of servers in the Consulting Rooms to be eight.

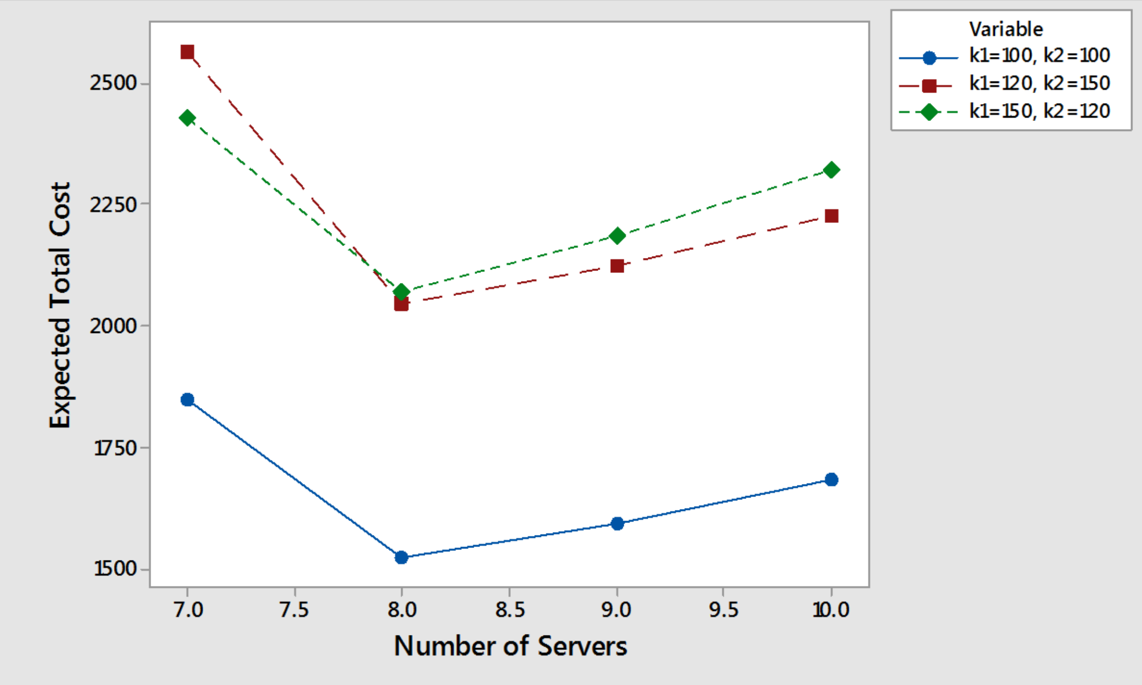

Fig. 5: Plot of Expected Total Cost against different numbers of Servers for different values of k1 and k2

Fig. 5: Shows Expected Total Cost Against Different Number of Servers, for Different Values of k_1 and k_2.

CONCLUSION

The study has revealed that there are high traffic intensities at the different units of the MOPD. Moving from the Card Room to the Vital Signs Unit, the traffic intensity increases; but the waiting time of the patients (in the queue and in the system) are not long, compared to what obtains at the Card Room. This is because of the presence of two servers, working as a LND3/LLD3/2 model.

However, at the Consulting Rooms where there are 7 servers, working as parallel single-queue-single-server channels, it is realized that the mean waiting times of patients are alarmingly long.

In a bid to make a trade-off between the increased costs of providing better service and decreased waiting time cost of customers at the Consulting Rooms, it is suggested that the system could operate optimally if a server were added to the unit to bring it to 8, hence, operating as a LLD3/G/8 model.

REFERENCES

Abdul-Wahab N. Y. and Najim U. (2014): An Application of Queueing Theory to ATM Service Optimization; A case study. *Mathematical Theory and Modelling*, 4(6): 11 – 23.

Adeleke R. A, Ogunwale O. D. and Halid O. Y. (2009): Application of Queueing Theory to Waiting-time of Out-Patients in Hospitals. *Pacific Journal of Science and Technology*, 10(2): 270 – 274.

Bhat U. N. (2008): *An Introduction to Queueing Theory; Modeling and Analysis in Applications*. New York. Birkhauser Boston Publisher.

Erlang A. K. (1909): The Theory of Probability and Telephone Conversations. *Nyt Tidsskrift Matematika*, 20(B): 33-39.

Gross D., Shortle J. F., Thompson J. M. and Harris C. (2008): *Fundamentals of Queueing Theory*. Hoboken, New Jersey. John Wiley and Sons, Inc.

Igwe A., Onwumere J. U. J. and Egbo O. P. (2014): Efficient Queue Management in Supermarkets; A Case Study of Makurdi Town, Nigeria. *European Journal of Business and Management*, 6(39): 185 – 192.

Jhala N. and Bhathawala P. (2016): Application of Queueing Theory in Banking Sector. IOSR Journal of Mathematics, 12(2): 73 – 75.

Jhala N. and Bhathawala P. (2017): Analysis and Application of Queueing Theory in Supermarkets. *International Journal of Innovative Research in Science, Engineering and Technology*, 6(9): 17974 – 17979.

Kembe, M. M, Onah, E. S. and Iorkegh, S. (2012): A Study of Waiting and Service Costs of a Multiserver Queueing Model in a Specialist Hospital. *International Journal of Scientific and Technology Research*, 1(8): 19 – 23.

Little J. D. C. (1961): A Proof for the Queueing Formula: $L = \lambda W$. *Operations Research*, 9(3): 383 -387.

Sharma J. K. (2010): *Operations Research; Theory and Applications*. Daryaganj, New Delhi. Macmillan Publishers India Ltd.