Queue Management Benchmark for Modelling Patients' Flow in Nigerian Public Hospitals

Akpan, Anietie Peter¹ and John, Efiok Nsikan²

¹Department of Business Management, University of Uyo, Uyo, Nigeria.
²Department of Operations and Supply Chain Management, Garden City Premier Business School, Port-Harcourt, Nigeria.

Authors’ contributions
This work was carried out in collaboration between both authors. Both authors read and approved the final manuscript.

Article Information
DOI: 10.9734/JEMT/2021/v27i83035
Editor(s):
(1) Dr. Kamarulzaman Ab. Aziz, Multimedia University, Malaysia.
Reviewers:
(1) Ali Gul, Pakistan.
(2) M. M. Rudi Sarwono, Institut Teknologi Dan Bisnis Ahmad Dahlan Jakarta, Indonesia.
Complete Peer review History: https://www.sdiarticle4.com/review-history/74678

ABSTRACT

Although queue management in hospitals is widely researched, little is known about the benchmark for modelling patients flow in terms of the optimal number of servers required for effective service delivery. This study applied the queuing theory to the Nigerian public hospitals by setting a benchmark for modelling patients flow. A mixture of survey and observation was adopted to garner data for 30 days from patients in six public hospitals in Nigeria. Data were subjected to performance analysis via the Temporary Ordered Routine Algorithm. The computed performance values were further compared with their acceptable benchmarks for multi-server queues through the General Purpose System Simulator. We found the queuing system in the select hospitals not in congruence with the system performance benchmark; the mean service rate in each facility was low compared to the mean arrival rate; and the simulated number of doctors for were below the modelled benchmark. Managerial implications of findings were discussed.

Keywords: Queuing system; queuing theory; queue management in hospital; server efficiency; temporary ordered routine algorithm.

*Corresponding author: E-mail: nsikiman5000@gmail.com;
1. INTRODUCTION

Queues are formed when the demand for service exceeds its supply thereby causing people or objects to wait for the supply of service or when the service facilities stand idle and wait for customers [1]. This implies that a queue is formed when the total number of customers requiring service exceeds the number of available and functioning service facilities, or when the total number of service facilities exceeds the number of customers requiring services. Waiting time depends on the number of customers -which could be human beings or objects- on the queue, the number of servers serving the queue, and the duration of service time for each customer [2].

Waiting in a queue is a common phenomenon around hospitals in developing nations including Nigeria. The situation is even worse in public hospitals where patients wait for long periods without accessing medical services [36]. When considering improvements in services, the healthcare manager weighs the cost of providing a given level of service against the potential costs arising from having patients waiting [3,36]. Furthermore, the manager of the system that provides service is concerned both with the adequacy of service rendered and with the efficient usage of available service facility. Queuing theory is used to evaluate the performance of an existing queuing system by assessing the system performance indicators and comparing these indicators with set benchmarks. Also, it helps the manager to optimise output at a minimum total cost to the system by examining the extent to which server scheduling matches the pattern of patients’ arrival into the queuing system

In hospitals, the rising cost of health care can be attributed not only to ageing population and new expensive and advanced treatment modalities but also to inefficiencies in healthcare delivery [34]. Queuing theory application in hospitals is an attempt to minimize the cost of service delivery through minimization of inefficiencies and delays in the service system. Ezirim & Nwokah [4] and Gombolay, Golen, Shah, & Shah, [5] are of the opinion that the major motivation for contemporary study and application of queuing theory is in finding a solution to the practical problem of queuing while waiting for service. Another motivation is in managing available resources to the extent that queuing is minimised and optimum results attained.

The management of queues in hospitals is rapidly gaining interest amongst scholars and practitioners, particularly because it is very essential in effective healthcare delivery [6,2]. There are many problems in health care system which can be solved using queuing theory in the hospital. These include estimation of waiting and service time in the system, patients flow, emergency room arrivals, hospital pharmacy and pharmacy stores, public health, and so on. However, empirical evidence is lacking, especially in the developing countries, on the application of queuing theory in the evaluation of queuing system by applying the queue management benchmark. In addition, while some past studies found that queuing theory aids effective queue management in hospitals [7,8], others reported in the contrary [9,10]. Moreover, much as many studies have indicated a growing application of queuing theory on the management of queues in hospitals, attention has been focused primarily on western economies [11,12]. A common feature amongst available studies is the large concentration on the general effect of waiting time and service time on queues by identifying the statistical distribution of the observed queues [13,14]. Little is known about queue management benchmark for modelling patients flow and queue optimality in terms of the number of servers necessary for effective service at the hospital patient waiting lines. The aim of this study is to contribute to the body of knowledge by applying queuing theory in the Nigerian hospital environment, with the view towards setting a benchmark for modelling patients flow in hospital queuing system.

1.1 Literature Review

The success or failure of any model is determined by its performance measures [15,16]. Performance measures and benchmarks are used to gain useful information about queue systems. They serve as the feedback loop of any planning cycle. Performance benchmarks for queue management include: The average waiting time, the traffic intensity coefficient which measure average utilisation of the queuing system, the Po coefficient indicating the probability that a system is idle, and the system utilisation rate-an indication of capacity utilisation. According to Nehmias [17], purely random arrivals and purely random service process in queuing, means the arrival and service time follow the Poisson and exponential process respectively. Furthermore, if arrivals follow a Poisson process, it means that inter-
arrival times have the exponential distribution, and because of the memory-less property of the exponential distribution, the Poisson process is referred to as a purely random arrival process.

Carter, [15] observes that the operating characteristic of queuing system are determined largely by two statistical properties, the probability distribution of inter–arrival time and the probability distribution of service time. For total queuing systems, these distributions can take on almost any form (with restriction to negative values). A useful model should provide reasonable predictions while at the same time, being sufficiently simple that the model is mathematically tractable. Based on these conditions, the exponential distribution seems to be the most important probability distribution for the study of queuing theory [35]. The exponential distribution plays an important role in queuing process. This is due to the no memory property which is used in modelling the inter–arrival times. This implies that to predict future arrival patterns, we need not keep track of how long it has been since the last arrival.

The memoryless property makes the Poisson and exponential distributions suitable for the analysis of queuing systems [18]. This property implies that the length of the time interval from the current time to the occurrence of the next event does not depend upon the time of occurrence of the last event. In the Poisson probability distribution, the observer records the number of events that occur in a time interval of fixed length. In the (negative) exponential probability distribution, the observer records the length of the time interval between consecutive events. In both cases, the underlying physical process is memoryless [19,20]. It is mainly because of this special property that the exponential distribution has been the most widely used distribution in the analysis of queuing systems. Models based on the Poisson process often respond to inputs from the environment in a manner that mimics the response of the system being modelled. The analytically tractable models that result yield both information about the system being modelled and the form of their solution [21,14].

There are ample empirical evidences of the queue philosophy being applied in modern healthcare management. For instance, Agnihotri & Taylor [7] examined ways to staffing a centralised appointment scheduling in Lourdes hospital. The study adopted an experimental research design by grouping periods that receive similar call intensity and determining the necessary staffing for each of such intensity. The study was carried out for a period of 60 days. Data collected for the study were analysed using Gentel’s model and scenario simulation were applied in line with the study of Xia, Sean, & Bruce, (2018). The finding of the study revealed that staffing varied dynamically with call intensity and that redistributing server capacity over time led to a reduction in queue length and waiting time. The study recommended that managers in a similar situation should vary staff schedule according to need and/or queuing tendency. Similarly, Stuart & Stuart [12] conducted a study on patients’ inter-arrival time and mean waiting time at the University of Iowa Teaching Hospital. They found that patient’s inter-arrival time and mean waiting time followed the exponential distribution and that waiting time could be reduced by a more efficient use of the servers. They recommended that servers should be put to efficient use by allocating 23 doctors per shift of not more than five (5) hours.

Brian [22] conducted a study on the effect of queue on motivation and performance outcomes of medical staff. The study sampled 4564 employees of both private and public hospitals in Italy. The study adopted a descriptive survey design and multi-stage sampling technique was used in selecting the sample. The Pearson Product Moment Correlation coefficient, multiple regression analysis and the independent t-test were used to analyse the data and the results obtained showed that employees were not motivated by long queues. Also, the study showed that there was a significant relationship between long queues and performance outcomes in hospitals. He, therefore, recommended that management of hospitals should work to improve performance so as to reduce queue length and waiting time.

Denver [11] studied the causes of long queues in select hospitals in the state of Rome, Italy. The study was a cross sectional descriptive study. A sample of six hospitals from the study area was selected using the stratified random sample and 642 respondents were randomly drawn from patients attending the out-patient departments of the chosen hospitals. The study revealed that inadequate facilities and personnel caused long waiting time in public hospitals in South Africa. It recommended that more facilities should be made available and more personnel should be employed to reduce the waiting time of patients attending hospitals.
A common feature of the reviewed empirical works is that they have concentrated mostly on either estimating inter-arrival time and mean waiting time, or modelling particular queue situations. Also, all the studies reviewed were done in the developed countries of Europe, North America and some Asian countries like India. None was done in a developing country like Nigeria. Moreover, none evaluated the application of the system performance benchmark in the general out-patient department of hospitals nor studied the effect of resource scheduling on queue and how it matches the arrival pattern of patients into the queue which this study seeks to contribute to the literature.

On the basis of the literature so far reviewed, the following hypotheses are formulated:

H₀₁: Tertiary hospitals in South-South Nigeria are not likely to meet the system performance benchmark for modelling patient flow in their general out-patient departments.

H₀₂: Server scheduling does not match the pattern of patients’ arrival in select tertiary hospitals.

2. METHODOLOGY

The survey research design was adopted in this study because it provides a quick, efficient and accurate means of assessing information about a population [23]. The target population in this case was all the patients that visits the General Outpatient Departments (GOPD) of all public tertiary hospitals in the South-South geopolitical zone of Nigeria for medical treatment.

There are eight public tertiary hospitals in the zone, but six federal tertiary hospitals were selected on the basis of one federal tertiary hospital in each state that makes up the South-South geopolitical zone. This approach was taken to eliminate bias and allow for a good framework for sampling. The framework for determining the sample size for this study was the average weekly population of patients that visit the GOPD of each of the select tertiary hospitals. Available records in the respective hospitals as at the time of study were used to compute the average weekly population of patients visiting the GOPD of these hospitals.

A three item questionnaire was self-designed and used to collect data on patients’ frequency of visit, queue experience, and waiting time experience in each hospital. Before data collection, the instrument was validated by two research experts and professionals in the GOPD unit. Data were collected for a period of thirty (30) working days. In addition, two research assistants were used. They observed and took note of the patients’ arrival time into the queue and departure time from the queue. A patient was deemed to have departed from the queue when he or she has been attended to by a doctor. Data collected through this procedure was used to simulate queue experience for all the multi-server systems in each hospital under study.

Using the modified version of Hillier & Liberman’s queuing model by McKnisto, [25], we employed the following model for this study:

\[ (M/M/s): (∞/FCFS) \] Where:

- \( M \) = arrival rate (i.e. Poisson distribution);
- \( M \) = queue configuration (multiple server);
- \( s \) = number of servers;
- \( ∞ \) = system capacity (infinite); and
- FCFS = First-come-first-served.

For hospital queuing system, it is assumed that the arrivals follow a Poisson probability distribution at an average of \( \lambda \) patients per unit of time [31]. It is also assumed that they are served on a first-come-first-served basis by any of the doctors and that the system has an infinite capacity. Also, it is assumed that the service times are distributed exponentially, with an average of \( \mu \) customers (patients) per unit of time and number of servers (s) [32]. In order to compute the performance measures of the queuing systems observed in the select tertiary hospitals, the Temporary Ordered Routine Algorithm (TORA) was used. The performance measures included mean waiting time, traffic intensity coefficient, and \( P_0 \) coefficient. The computed values of these performance measures in each of the hospitals were then compared with their acceptable benchmarks for multi-server queuing systems through scenario simulation technique. This simulation was done using the General Purpose System Simulator (GPSS), a system software that was programmed to depict reality in a system [33].

3. RESULTS AND DISCUSSION

3.1 Results

H₀₁: Tertiary hospitals in South-South Nigeria are not likely to meet the system performance
benchmark for modelling patient flow in their general out-patient departments.

After subjecting the above hypothesis to test, the results are as summarised on Table 1. Table 1 presents the multi-server performance measure analysis for the six (6) select hospitals. It was observed that the mean waiting time for Federal Medical Centre, Yenegoa (FMCY) was 299 minutes while that of Federal Medical Centre, Asaba (FMCA) was 278 minutes. University of Calabar Teaching Hospital (UCTH) and University of Uyo Teaching Hospital (UUTH) posted mean waiting time of 266 and 253 minutes respectively while the University of Benin Teaching Hospital (UBTH) and the University of Port Harcourt Teaching Hospital (UPTH) had mean waiting time of 279 and 302 minutes respectively. These values are higher than the acceptable queue performance benchmark of 240 minutes set by the World Health Organisation for tertiary health facilities [26]. This implies that patients visiting all the select hospitals for medical treatment wait for more than 240 minutes (4 hours) before being attended to by a doctor.

For the traffic intensity coefficient, it was observed that FMCY posted 1.51, FMCA 1.39 and UCTH 1.42. UUTH was 1.33 while UBTH and UPTH had 1.15 and 1.41 respectively. The above traffic intensity coefficients are higher than the acceptable coefficient for an efficient multi-server queuing system which is 0.80 [26]. This high traffic intensity coefficient explains the long waiting time experienced in the general outpatient department of the six (6) hospitals.

Furthermore, the \( P_0 \) coefficient of 0.07, 0.16 and 0.19 were observed for FMCY, FMCA and UCTH respectively. Those for UUTH, UBTH and UPTH were 0.23, 0.20 and 0.19 respectively. These coefficients are below the acceptable benchmark of 0.30 [9]. Therefore, the null hypothesis is upheld. This implies that the select tertiary hospitals are not meeting the system performance benchmark for modelling patient flow in their general out-patient departments.

**H\(_2\):** Server scheduling does not match the pattern of patients’ arrival in select tertiary hospitals.

The result of the test of hypothesis 2 are presented on Tables 2, 3 and 4.

Table 2 shows the summary of queue simulation for a multi-server system in FMCY and FMCA. From the simulation result, it was observed that a unit increase in the number of doctors in both hospitals yielded significant reduction in the values of the expected waiting time in queue (\( T_q \)), expected number of patients waiting in queue (\( N_q \)), expected number of patients waiting in the system (\( N_s \)) and traffic intensity coefficient. Therefore, the null hypothesis is accepted. This implies that at the current level of 5 and 6 doctors at the general outpatient department of FMCY and FMCA respectively, server scheduling does not match the pattern of patients’ arrival in the two hospitals.

From the simulation result in Table 3, it was also observed that a unit increase in the number of doctors in UUTH and UCTH yields significant reduction in the values of the expected waiting time in queue (\( T_q \)), expected number of patients waiting in queue (\( N_q \)), expected number of patients waiting in the system (\( N_s \)) and traffic intensity coefficient. Therefore, the null hypothesis is accepted. This implies that at the current level of 7 doctors each at the general outpatient department of FMCY and FMCA respectively, server scheduling does not match the pattern of patients’ arrival in the two hospitals.

Furthermore, the simulation result in Table 4 shows also that a unit increase in the number of doctors in UBTH and UPTH yields significant reduction in the values of the expected waiting time in queue (\( T_q \)), expected number of patients waiting in queue (\( N_q \)), expected number of patients waiting in the system (\( N_s \)) and traffic intensity coefficient. Therefore, the null hypothesis is accepted. This implies that at the current level of 9 doctors each at the general

<table>
<thead>
<tr>
<th>Performance Measures</th>
<th>FMCY</th>
<th>FMCA</th>
<th>UCTH</th>
<th>UUTH</th>
<th>UBTH</th>
<th>UPTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean arrival rate (( \lambda ))</td>
<td>32</td>
<td>37</td>
<td>49</td>
<td>44</td>
<td>52</td>
<td>58</td>
</tr>
<tr>
<td>Mean service rate (( \mu ))</td>
<td>8</td>
<td>8</td>
<td>12</td>
<td>11</td>
<td>13</td>
<td>14</td>
</tr>
<tr>
<td>Number of service providers</td>
<td>5</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>Mean waiting time (in minutes)</td>
<td>299</td>
<td>278</td>
<td>266</td>
<td>253</td>
<td>279</td>
<td>302</td>
</tr>
<tr>
<td>Traffic intensity coefficient</td>
<td>1.51</td>
<td>1.39</td>
<td>1.42</td>
<td>1.33</td>
<td>1.15</td>
<td>1.41</td>
</tr>
<tr>
<td>( P_0 ) coefficient</td>
<td>0.07</td>
<td>0.16</td>
<td>0.19</td>
<td>0.23</td>
<td>0.20</td>
<td>0.19</td>
</tr>
</tbody>
</table>
Table 2. Summary of queue simulation for a multi-server system in FMCY and FMCA

<table>
<thead>
<tr>
<th>Performance Measures</th>
<th>FMCY</th>
<th>FMCA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5 Docs</td>
<td>6 Docs</td>
</tr>
<tr>
<td>Mean arrival rate ((\lambda))</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>Mean service rate ((\mu))</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>System utilisation (%)</td>
<td>89.8</td>
<td>88.1</td>
</tr>
<tr>
<td>Mean waiting time (mins)</td>
<td>291</td>
<td>263</td>
</tr>
<tr>
<td>Traffic intensity coeff.</td>
<td>1.51</td>
<td>1.01</td>
</tr>
<tr>
<td>Nq</td>
<td>20</td>
<td>18</td>
</tr>
<tr>
<td>Ns</td>
<td>44</td>
<td>37</td>
</tr>
<tr>
<td>Tq (mins)</td>
<td>258</td>
<td>241</td>
</tr>
<tr>
<td>Po coefficient</td>
<td>0.07</td>
<td>0.29</td>
</tr>
</tbody>
</table>

Table 3. Summary of queue simulation for a multi-server system for UCTH and UUTH

<table>
<thead>
<tr>
<th>Performance Measures</th>
<th>UCTH</th>
<th>UUTH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>7 Docs</td>
<td>8 Docs</td>
</tr>
<tr>
<td>Mean arrival rate ((\lambda))</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>Mean service rate ((\mu))</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>System utilisation (%)</td>
<td>89.8</td>
<td>87.1</td>
</tr>
<tr>
<td>Mean waiting time (mins)</td>
<td>266</td>
<td>248</td>
</tr>
<tr>
<td>Traffic intensity coeff.</td>
<td>1.42</td>
<td>1.27</td>
</tr>
<tr>
<td>Nq</td>
<td>21</td>
<td>16</td>
</tr>
<tr>
<td>Ns</td>
<td>56</td>
<td>51</td>
</tr>
<tr>
<td>Tq (mins)</td>
<td>237</td>
<td>229</td>
</tr>
<tr>
<td>Po coefficient</td>
<td>0.19</td>
<td>0.24</td>
</tr>
</tbody>
</table>
Table 4. Summary of queue simulation for a multi-server system for UBTH and UPTH

<table>
<thead>
<tr>
<th>Performance Measures</th>
<th>UBTH 9 Docs</th>
<th>UBTH 10 Docs</th>
<th>UBTH 11 Docs</th>
<th>UBTH 12 Docs</th>
<th>UBTH 13 Docs</th>
<th>UPTH 9 Docs</th>
<th>UPTH 10 Docs</th>
<th>UPTH 11 Docs</th>
<th>UPTH 12 Docs</th>
<th>UPTH 13 Docs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean arrival rate ($\lambda$)</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>52</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
<td>58</td>
</tr>
<tr>
<td>Mean service rate ($\mu$)</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>13</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>System utilisation (%)</td>
<td>88.2</td>
<td>87.6</td>
<td>83.3</td>
<td>71.1</td>
<td>62.0</td>
<td>90.3</td>
<td>89.4</td>
<td>87.6</td>
<td>84.1</td>
<td>69.4</td>
</tr>
<tr>
<td>Mean waiting time (mins)</td>
<td>279</td>
<td>260</td>
<td>233</td>
<td>221</td>
<td>214</td>
<td>302</td>
<td>281</td>
<td>264</td>
<td>239</td>
<td>229</td>
</tr>
<tr>
<td>Traffic intensity coeff.</td>
<td>1.15</td>
<td>1.07</td>
<td>0.78</td>
<td>0.63</td>
<td>0.57</td>
<td>1.41</td>
<td>1.13</td>
<td>0.92</td>
<td>0.77</td>
<td>0.52</td>
</tr>
<tr>
<td>Nq</td>
<td>24</td>
<td>19</td>
<td>14</td>
<td>11</td>
<td>8</td>
<td>30</td>
<td>26</td>
<td>21</td>
<td>14</td>
<td>12</td>
</tr>
<tr>
<td>Ns</td>
<td>61</td>
<td>56</td>
<td>50</td>
<td>27</td>
<td>25</td>
<td>73</td>
<td>70</td>
<td>64</td>
<td>58</td>
<td>41</td>
</tr>
<tr>
<td>Tq (mins)</td>
<td>261</td>
<td>247</td>
<td>232</td>
<td>201</td>
<td>187</td>
<td>272</td>
<td>261</td>
<td>250</td>
<td>239</td>
<td>163</td>
</tr>
<tr>
<td>Po coefficient</td>
<td>0.20</td>
<td>0.25</td>
<td>0.39</td>
<td>0.46</td>
<td>0.50</td>
<td>0.09</td>
<td>0.16</td>
<td>0.28</td>
<td>0.35</td>
<td>0.37</td>
</tr>
</tbody>
</table>
outpatient department of UBTH and UPTH, server scheduling does not match the pattern of patients’ arrival in the two hospitals.

Results of simulation revealed that the mean waiting time, expected waiting time in queue (Tq), expected number of patients waiting in queue (Nq), expected number of patients waiting in the system (Ns) and traffic intensity coefficient of the different queuing systems observed in the general outpatient department of the hospitals could be reduced when there is an increase in the number of doctors attending to patients in those hospitals. This indicates that the number of doctors currently on daily duty in the hospitals are not sufficient. This implies that server scheduling does not match the pattern of patients’ arrival in select tertiary hospitals.

3.2 Discussion

The purpose of this paper had been to investigate the effectiveness of queue management in public hospitals with the view towards setting a benchmark to model patients’ flow. The study reveals that the queuing system in each of the select hospitals was not meeting the system performance benchmark for modelling patient flow in their general out-patient department. This position was reached after computing the values of the waiting time, traffic intensity coefficient and the Po coefficient in Table 4. Taking traffic intensity, for instance, Komashie, Mousavi, Clarkson, & Young, [21] opines that a traffic intensity coefficient of above 0.80, yields longer waiting time by customers in a multi-server queuing system. This position corroborates Enim [18] and Kandemir-Caues & Cauas [1] that there is a significant inverse relationship between traffic intensity and waiting time in a multi-server queuing system. The only way that the traffic Intensity can be kept at a reasonable figure (i.e. below or equal to 0.80) is to provide adequate number of servers, assuming of course that the arrival rate cannot be controlled [27].

Also, Yaduvanshi, Sharma, & More, [28] opines that average waiting time in a tertiary hospital is considered excessive by the World Health Organisation (WHO) when it is above 240 minutes (4 hours). This implies that patients wait excessively in the select hospitals where average waiting time is found to be above 240 minutes. Majority of the respondents (86.3%) said that they wait in queue in the general outpatients’ department of the select hospitals for above 4 hours before being attended to by a doctor. According to Xia, Sean, & Bruce, [20], excessive waiting time in hospitals is an indication that the hospital is not meeting the system performance benchmark for modelling patient flow.

Furthermore, the finding of the study reveals that server scheduling does not match the pattern of patients’ arrival in the select tertiary hospitals. This position was arrived at through queuing system simulation. The result of the system simulation shows a lack of optimality in server scheduling in all the select hospitals. A closer look at the results reveal that the optimal server level for each of the select hospital is achieved when the number of doctors in the general outpatients’ department is increased to 7 for FMCY and FMCA, 9 for UCTH, 8 for UUTH and 11 and 12 for UBTH and UPTH respectively. This is so because at these points of optimality the mean waiting time at the general outpatients’ department for each of the select hospitals is almost at equilibrium with the expected waiting time (Tq). Beyond these points of optimality, it is noticed that further increment by a unit of doctor yielded a diminishing marginal return in the values of the expected numbers of patients waiting in queue (Nq) and the expected numbers of patients waiting in the system (Ns). This suggests a decline in marginal productivity when the number of doctors is increased beyond the points of optimality for each of the select hospitals. In the opinion of Kennedy, Rhodes, & Asplin, [29], the optimality of resource scheduling is reached at the point where there is equilibrium or near equilibrium (not more than 6-unit difference) between the prevailing mean waiting time and the expected waiting time. This is so because at this point the servers will be optimally busy and marginal productivity is maximised. Nosek & Wilson [30] opines that this point balances the trade-off between cost of patients waiting and the cost of providing service.

4. CONCLUSION AND IMPLICATIONS

Although queue management is widely discussed in management literature, a model to evaluate queue management benchmark for modelling patients’ flow in the general outpatients’ department in public hospitals has not been sufficiently specified. Thus, this study advances the field of queue management by showing that effectiveness and efficiency can be achieved in management of queues in tertiary hospitals through the application of queuing theory. It is concluded that the queuing systems
in the select tertiary hospitals were not meeting the system performance benchmark for modelling patient flow in their general out-patient departments. This means that patients waited for long hours before being attended to by a doctor. The performance measure as used in the study included mean waiting time, traffic intensity coefficient and Po coefficient. The study also revealed that the mean service rate in each of the hospitals was low compared to the mean arrival rate. This means that instead of decreasing, the number of patients waiting in the system will continue to increase. The study further revealed that server scheduling did not match the pattern of patients’ arrival in select tertiary hospitals. This means that the number of doctors on duty per time at general outpatients’ department of the select hospitals was not enough to produce a service rate that would reduce the patients’ waiting time to the acceptable limit. Queuing system simulation results revealed that the optimal number of doctors for FMCY and FMCA was 7 doctors each per duty time. That of UUTH was 8 doctors, UCTH 9 doctors while UBTH and UPTH needed 11 and 12 doctors respectively to achieve optimality.

Based on the findings, it is concluded that the queuing system in the select tertiary hospitals are not meeting the system performance benchmark for modelling patient flow in their general out-patient departments. Also, server scheduling does not match the pattern of patients’ arrival in the select tertiary hospitals. Hospitals that are mindful of these performance measures are likely to reduce patients waiting time in their general outpatients’ departments to the acceptable limit thereby reducing the queue length in those hospitals.

In terms of policy implications for public hospitals, a compelling case can be made from these findings for scheduling the optimal number of doctors per shift in GOPDs in order reduce in-patients’ waiting time and enhance their satisfaction. In addition, more doctors should be recruited to enhance balance between cost of providing service and the cost of waiting. Since patients’ waiting time in the general outpatients’ departments were found to be excessive, hospital management should look into other ways of enhancing patients’ waiting time experience such as providing a wide variety of current magazines and newspapers for the patients who are waiting, and GOPDs should be equipped with television sets to further improve patients’ waiting experiences.

**DISCLAIMER**

The products used for this research are commonly and predominantly use products in our area of research and country. There is absolutely no conflict of interest between the authors and producers of the products because we do not intend to use these products as an avenue for any litigation but for the advancement of knowledge. Also, the research was not funded by the producing company rather it was funded by personal efforts of the authors.

**CONSENT**

As per international standard or university standard, patients’ written consent has been collected and preserved by the author(s).

**ETHICAL APPROVAL**

As per international standard or university standard written ethical approval has been collected and preserved by the author(s).

**COMPETING INTERESTS**

Authors have declared that no competing interests exist.

**REFERENCES**

5. Gombolay M, Golen T, Shah N, Shah J. Queueing theoretic analysis of labor and


