Improving the Efficiency of Outpatient Services at Benue State University Teaching Hospital using the Queuing Theory

Ishaku Ara Bako¹*, Priscilla M. Utoo¹ and Jonathan Ikughur²

¹Department of Epidemiology and Community Health, College of Health Sciences, Benue State University, Makurdi, Nigeria
²Department of Mathematics/Statistics/Computer Science, University of Agriculture, Makurdi, Benue State, Nigeria

Abstract: Introduction: Long client waiting time is a characteristic of poor performance of the health care delivery and is a major challenge for healthcare services all over the world, especially in developing countries. The study was aimed at developing a model that optimizes performance of the general outpatient department of the Benue State University Teaching Hospital, Makurdi, Benue State Nigeria.

Methodology: Data was collected through observation and interviews with doctors at the general outpatient clinic of the Benue State University Teaching Hospital. The average number of clients seen per day was calculated by determining the average of daily attendants for five consecutive working days. The data obtained was used to create a five capacity scenarios using the queuing theory software.

Result: The Average Daily Attendance (ADA) was 73.2 clients while the Average Daily Arrival Rate was 10.47 clients per hour. There were six doctors working on any given day in the clinic and a doctor spends an average of 16.2 minutes per patient, representing an average of 3.7 patients per hour. The model showed that the optimum system performance can be achieved with four doctors (with 70.7% server utilization rate, average of 1.065 clients on the queue and 0.102 hours waiting time).

Conclusion: Four doctors working at the same time at the general outpatient clinic is required for optimal performance. The queuing theory should be used regularly at GOPD BSUTH and in all health facilities experiencing long queues to optimize operational efficiency.

Keywords: Waiting time, Client satisfaction, server, performance, Makurdi.

1. INTRODUCTION

1.1. Background

The expectations of the clients of health services not necessarily be in line with the providers’ priorities. A patient centered approach in health care is therefore increasingly being advocated by the World Health Organization (WHO) [1]. This entails that services are sensitive to patients’ cultural values and expectations [2]. This has become more imperative because in an emerging competitive health care market, health care managers must aim at obtaining excellent rating from patients or clients in order to continue to attract them [3].

It is the desire of health care managers to increase the number of clients/patients attending the facilities. However, this is often at the cost of long waiting time, compromised standards and low clients’ satisfaction [4]. The cost involved in providing health care service, among many others, include, personnel cost especially salaries, cost of building equipment, instruments and supplies. The relationship between number of patients attending a facility and client satisfaction is a virtuous cycle. Health facilities are often faced with the challenge of limited resources and the cost of human resource is increasing in the midst of dwindling funding.

It is the desire of the average health care manager to ensure that servers are busy rather than idle. However, maximal utilization of the servers will be at the cost of the length of the service line and the average waiting time, and as the server utilization rate increases, the longer the service line and the patient waiting time. Queuing model has been applied in many settings including health care to maximize the use of available resources and optimize waiting time for clients.

Patient waiting time has been found to affect clients’ satisfaction with services. A research on client satisfaction with primary care services in USA showed that longer waiting time was associated with lower client satisfaction but increased time spent with the health care worker increased level of satisfaction [6]. In a client satisfaction survey conducted at the outpatient department (OPD) of the National Hospital Abuja, Nigeria, there was a statistically significant association between waiting time and patient satisfaction with services [7].
Queuing theory is a mathematical approach which has been applied as a management decision making tool to improve waiting time and health system performance. The concept was first described by Erlang AK in 1913 [8] (and is aimed at matching demand for services with optimal resources utilization, minimal client waiting time and by extension maximal client satisfaction. Several studies have shown that the queuing theory can be effectively applied to the health care settings. It was successfully used to optimize the management of an outpatient pharmacy of a military health facility in Iran. The study showed that increasing one staff in the unit could reduce the average queue length by 10 and reduce average waiting time by 18 minutes 14 seconds [9].

The study was aimed at developing a model for measuring and improving waiting time among patients/clients and operational efficiency of the general outpatient department of the Benue State University Teaching Hospital (GOPD BSUTH), Makurdi, Benue State Nigeria.

2. METHODOLOGY

2.1. Study Setting

The study was conducted at the general outpatient department of the Benue State University Teaching Hospital, Makurdi, Benue State. The hospital provides tertiary health care services for citizens of Benue state and surrounding states including Nasarawa, Taraba and Cross River states. The general outpatient department was one of the first set of clinics to be established in the hospital and is manned by the Department of Family Medicine of the hospital. The clinic is the 'clearing house' of the hospital and accordingly, is the first to see non-emergency cases coming to the hospital.

2.2. Study Population

The study was conducted among clients attending the general outpatient department of the Benue State University Teaching Hospital, Makurdi. From the inception of the hospital, the number of clients attending the clinics in GOPD has increased from 6504 in 2012 to 19778 in 2015.

2.3. Data Collection Procedure

Data was collected during working hours for five days (Monday to Friday). Observation technique was used to observe and document the time when patients or clients arrive at the outpatient department and a time they come out from outpatient department after receiving the required services. Interviews were also conducted with the doctors attending to patients at the clinic.

2.4. Statistical Data Analysis

Assumptions in Queuing Theory

a. The arrival of customers into the system is discrete and follows Poisson distribution with arrival rate, λ
b. The queuing discipline is first come first serve.
c. The service channel can only render service of finite rate exponentially distribution with service rate, μ.
d. The number of customer in the system at the time of initiating the observation is assumed to have arrived in the first unit time.

The Following Computations were Performed:

i. Average Daily Attendance (ADA), λ: = Average daily attendance at the clinic was determined using both secondary and primary data from attendance at the GOPD BSUTH previous 5 days as follows:

\[ \bar{X} = \frac{\sum x}{n} \]

Where \( \sum x \) = The sum of daily attendance for five consecutive working days

\( n = \) number of days observed = 5 and

\( \bar{X} = \) The average daily attendance (ADA)

ii. Service Rate, μ: The average number of customers served per unit of time and it is determined by estimating the average time a doctor takes to see a patient and then average number of patients seen by one doctor in one hour.

iii. Number of servers (doctors), s: 3, 4, 5.6 and 7

Performance Measures

The parameters were determined using online MMS software at the following website: http://people.revoledu.com/kardi/tutorial/Queuing/MMs-Queuing-System.html#MMsQueuingSystem [10]
Table 1: Shows a Breakdown of the Performance Measures and how they are Calculated [11]

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Formula for M/M/s</th>
</tr>
</thead>
<tbody>
<tr>
<td>Utilization factor (% of time the server is being utilized by a patient)</td>
<td>$\rho = \lambda / s\mu$</td>
</tr>
<tr>
<td>Probability that there are no customers in the system</td>
<td>$P_0 = \sum_{i=0}^{\infty} \left( \frac{\rho^i}{i!} \cdot \left( \frac{s\mu}{\mu+s\mu+\lambda} \right)^i \cdot \left( 1 - \frac{\lambda}{\mu} \right) \right)$</td>
</tr>
<tr>
<td>Probability that there are no customers in the system</td>
<td>$P_0 = 1 - (\lambda / \mu)^s$</td>
</tr>
<tr>
<td>Average number of patients in the queue for service</td>
<td>$L_q = \frac{P_0\rho}{s} + \frac{1}{s(1-\rho)}(s-p)^s$</td>
</tr>
<tr>
<td>Average number of patients in the system</td>
<td>$L_s = \frac{\rho s}{\mu} + \frac{1}{s(1-\rho)}(s-p)^s$</td>
</tr>
<tr>
<td>Average waiting time a patients spends waiting in line excluding the service time</td>
<td>$W_q = \frac{L_q}{\lambda}$</td>
</tr>
<tr>
<td>The average time a patient spends in the system (in waiting line and being served)</td>
<td>$W_s = W_q + \frac{1}{\mu} = L_s / \lambda$</td>
</tr>
</tbody>
</table>

3. RESULTS

3.1. Observations on the Queuing System and Service Discipline by Study Team

Our observation in the GOPD revealed that the multiple Single-server queuing model was employed i.e. patients on arrival to the clinic join a queue to see any of the average of about 6 doctors in the clinic consulting at the same time. Every client on arrival is seen by the nurse who took vital signs and then directs them to join any of the queues to see the doctor. Each doctor therefore has queue of patients waiting to be attended. After been seen by the doctor, they depart the clinic to either go home or be referred to other units of the hospital such as Pharmacy or the laboratory. The Queuing discipline is first come, first to be served except.

3.1.1. Input parameters

i. Average Daily Attendance (ADA)

Therefore: Average Daily Attendance (ADA) = (59+76+91+71+69)/5 = 73.2

ii. Average Daily Arrival Rate

The general outpatient clinic opens from 8:00 am till 4.00pm, with one hour break, giving a total of seven (7) hours per day. Using this information, the average patient arrival rate is given as total number of patients per day/total hours the clinic is opened

= 73.2/7

= 10.47 patients per hour

iii. Service Rate

From interviews with the doctors, the following represent the average time spent on each patient:

One doctor spends an average of 16.2 minutes per patient. Therefore, a doctor can see on average 3.7 patients per hour.

iv. Service Capacity

A total of one consultant and five resident doctors attend to patient daily at the GOPD of BSUTH. Therefore there are six servers.

3.1.2. Result of The Queuing Model

Five scenarios of server capacity were considered, comparing capacity utilization, average number in the queue, average number in the system, average waiting time in the queue and average waiting time in the system.

Table 2: Number of Patients Seen in Previous Five Days at the GOPD BSUTH Makurdi, November 2016

<table>
<thead>
<tr>
<th>SN</th>
<th>Day</th>
<th>Date</th>
<th>Number of patients</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>Thursday</td>
<td>3/11/2016</td>
<td>59</td>
</tr>
<tr>
<td>ii</td>
<td>Friday</td>
<td>4/11/2016</td>
<td>76</td>
</tr>
<tr>
<td>iii</td>
<td>Monday</td>
<td>7/11/2016</td>
<td>91</td>
</tr>
<tr>
<td>iv</td>
<td>Tuesday</td>
<td>8/11/2016</td>
<td>71</td>
</tr>
<tr>
<td>v</td>
<td>Wednesday</td>
<td>9/11/2016</td>
<td>69</td>
</tr>
</tbody>
</table>
Table 3: Average Time Spent per patient by the doctors at the GOPD BSUTH Makurdi, November 2016

<table>
<thead>
<tr>
<th>SN</th>
<th>Time per patient (minutes)</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>12</td>
</tr>
<tr>
<td>ii</td>
<td>12</td>
</tr>
<tr>
<td>iii</td>
<td>20</td>
</tr>
<tr>
<td>iv</td>
<td>15</td>
</tr>
<tr>
<td>v</td>
<td>25</td>
</tr>
<tr>
<td>vi</td>
<td>13</td>
</tr>
<tr>
<td>Overall Average time per patient</td>
<td>16.2</td>
</tr>
</tbody>
</table>

4. DISCUSSION

The analysis reveals that there are six doctors attending to clients per day and clients are attended to within seven (7) hours on working days. The average arrival rate and service rate of 10.47 patients per hour and 3.7 patients per hours respectively. Generally, server utilization reduces with increasing number of servers, the highest been 94.3% with 2 servers and lowest, 40.4% for 7 servers. The average time waiting in queue showed a corresponding downward trend, with the highest time spent been 1.42 hours for 3 servers and the lowest, 0.002 hours (implying almost no queuing) for 7 servers.

Server utilization was 70.74% and waiting time was 0.102 hours (6.12 minutes) for four (4) servers. This represents a change from 1.42 hours (1 hours, 25 minutes) for 3 servers which is huge. On the other hand, time spent in queue when 6 servers are used, as it is currently being done, is 0.007 hours with server utilization of 47.162%. Studies have established that below a certain server threshold, the queue length and waiting time for clients becomes very long [12].

Table 4: Output of the M/M/s Queuing Modelling of the GOPD BSUTH Makurdi, November 2016

<table>
<thead>
<tr>
<th>No. servers (Scenarios)</th>
<th>Average number of customers in system</th>
<th>Average number of customers in Queue</th>
<th>Average time spent in system (Hrs)</th>
<th>Ave time waiting in queue (Hrs)</th>
<th>Server utilization (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>Ls</td>
<td>Lq</td>
<td>Ws</td>
<td>Wq</td>
<td>p</td>
</tr>
<tr>
<td>3</td>
<td>17.698</td>
<td>14.869</td>
<td>1.690</td>
<td>1.420</td>
<td>94.324</td>
</tr>
<tr>
<td>4</td>
<td>3.894</td>
<td>1.065</td>
<td>0.372</td>
<td>0.102</td>
<td>70.743</td>
</tr>
<tr>
<td>5</td>
<td>3.085</td>
<td>0.256</td>
<td>0.295</td>
<td>0.024</td>
<td>56.595</td>
</tr>
<tr>
<td>6</td>
<td>2.900</td>
<td>0.070</td>
<td>0.277</td>
<td>0.007</td>
<td>47.162</td>
</tr>
<tr>
<td>7</td>
<td>2.849</td>
<td>0.019</td>
<td>0.272</td>
<td>0.002</td>
<td>40.425</td>
</tr>
</tbody>
</table>

The first scenario with four servers seems like the threshold, below which we had a huge increase in client waiting time. It can be observed from the finding that even though server utilization decreases with increasing number of servers, there was no significant change in waiting time from server number 4 and above. This trend is also true for number of clients in the queue, average number of customers the system and probability idle. Our results are in agreement with previous studies which showed that increasing the server was associated with increased performance measures: waiting time and queue length [12, 13]. It is worthy of note that with every additional server, there is a corresponding increase in cost to the hospital in terms of personnel cost, cost of equipment and consultation room construction and maintenance. A system is said to be performing optimally if it is both effective and efficient i.e. providing good quality service at minimal cost. In view of the above, the general outpatient department of BSUTH can deliver most efficiently with four doctors manning the clinic instead of the current 6 doctors.

Before now, most health care managers do not have enough evidence to use to decide on actions to take in managing queues. The balance between satisfying clients’ needs especially as it relates to waiting time and the need to curtail operational cost is very delicate. The Study has however confirmed that the queuing theory can be used as a management decision tool for the general outpatient department of the Benue State University Teaching hospital. The modeling exercise requires little data and so is relatively simple and fast to apply. The major limitation of the model is that the theory assumes a steady state meaning that for each number of customers in the system, the transition rate into the system is balanced by exit from the system. This condition assumes that
the queuing system is opened 24 hours a day and 7 days a week, but in reality, this is not achievable.

5. CONCLUSION

The study showed that the average arrival rate was 10.47 clients per hour and service rate was 3.7 clients per hour. Different server scenarios were used to determine the queuing performance. All the performance measures (except server utilization) improved with increasing number of servers. However, when compared with server utilization rate, four servers lead to the best combination of server utilization, number in queue and waiting time in queue. Four servers will be the optimal to achieve operational efficiency and good client waiting time and queue length given the current patient flow at the general outpatient department of Benue State University, Makurdi.

The application of queuing theory should be a continuous process. It is the desire of the hospital to have increase in the number of clients receiving services at the hospital. It is therefore, envisaged that the inputs used for this modeling will continue to change, necessitating regular repeat of the exercise. There is also the need for further queuing theory modeling that will include the cost of providing services. In our analysis, it was assumed that the higher the number of servers, the more the cost involved but the hospital administration would be better informed if details of the costing are worked out as well.

REFERENCES