An Improved Queuing Model for Reducing Average Waiting Time of Emergency Surgical Patients Using Preemptive Priority Scheduling

Shaista Rais[1], Dr. Humera Tariq[2], Dr. Shabnam Rais[3], Dr. Shakeel Mansoori[4]
[1]University of Karachi, Pakistan, shaista.rais@uok.edu.pk
[2]University of Karachi, Pakistan, humera@uok.edu.pk
[3]Karachi Medical & Dental College, Pakistan, shabnam.rais@gmail.com
[4]Karachi Medical & Dental College, Pakistan, drshakeelmansoori@yahoo.com

ABSTRACT

Our research objective is to reduce the Average Waiting Time for patients in an Emergency Department of public sector hospital. We have based our model on M/M/s Queuing System, our study reveals significant findings on arrival rate of patients. During this simulation, we have used a preemptive priority scheduling model. In our practice, the arrival rate followed a Poisson distribution, averaging 30 patients per hour, with the Mean Service time of 1.5 hours and Average Waiting Time recorded around 12.13 minutes. This research offers valuable help to achieve better time management in emergency departments of high-density medical facilities.

Keywords: Queuing, Inter-arrival Time, Service time, Emergency Surgical Patients, Simulation, Chi-Square Goodness of Fit Test, M/M/s Queuing Model, Preemptive Priority Scheduling.

1. INTRODUCTION

Emergency Department (E.D) or Trauma Centre is a critical component of any Health Centre facility and is often a site of crowding and chaos, at times jeopardising precious human lives. There have been numerous analytical studies to assess the form and extent of this crowding in the last few decades [1, 2] to develop better management practices and possibly to automat techniques to expedite these operations and procedures.

We also have embarked on a similar study to assess emergency surgical patients' Interarrival time and Service time distributions. In this pursuit, we have surveyed one of the busiest Trauma Centre (knows as the Emergency Department of Abbasi Shaheed Hospital), a well-known public sector hospital in Karachi, the largest city of Pakistan with around 50 million residents. To assess the Service time distribution in discrete timing, researchers have employed some techniques, both parametric [3] and non-parametric [4]. Taking a step ahead, using rigorous mathematical analysis, Schweer and Wichelhaus [5] have attempted to estimate the Cumulative Distribution Function, CDF [6], of the Service time distribution, based solely on the information of the arrival and departure processes. They devised a Service time distribution G in a discrete-time GI /G model, which is a queuing model spanning a space with an infinite number of servers. They used a functional form of the Central Limit Theorem [7]. They found mathematical proof for this theorem aided by some tenets of chaos theory, indicating the possibility of Service time estimation using the available arrival and departure processes data.

Computer simulation of operations in healthcare facilities and emergency departments, using various tools and techniques, has been a topic of extensive research in recent years [8, 9]. For a detailed review in this context, a treatise is present in literature [10]. Using various models, investigators have studied the patient arrivals in diagnostic facilities, such as radiology departments [11, 12]. In terms of service with a broader scope, an extensive study by Bhattacharjee and Ray [13] is available, where they studied and presented a detailed treatise on healthcare delivery processes in medical centres.

During the last century, to understand healthcare facilities, their various processes, and the essential concepts of patient flow, a necessary "Queuing theory" technique was developed [14,15]. Since then, extensive work has been carried out in this field of study to find optimised solutions for healthcare facilities using queuing analysis. In this context, a report by Zonderland and Boucherie [16] provides a valuable literature review. It gives an account of two examples where "Queuing Theory" was implemented to analyse healthcare facility using a queuing network. After reviewing the literature of different authors [17-19], we have realised that service timealways represents hospitals' overall performance, including all kinds of resources required to provide timely medical care to patients. Based on all these studies, we surmised that the service time seems to follow Gamma, Normal and Exponential Distributions.

Patient flow is difficult to understand due to patients' arrival time and service time [18]. Arrival time and Service time are those events that involve uncertainty in nature. Due to the

randomness and unpredictability in arrival and service processes, probability and statistics are utilised by queuing models and procedures [17]. In these contexts, the queuing theory and simulation analytic techniques are applied with acceleration as valuable means. It defines patients' arrival time and service time coming to the trauma centre with the proper distribution. The basic information of these models is the samples of arrival and Service time. These patterns are usually described by appropriate random distributions [18].

However, in the present study, patients' Interarrival time is exponentially distributed. After we observed the arrival process about how the events or individuals reach a certain point for a service, we have understood that the facility refers to urgent medical care for which patients arrive at the hospital. Nevertheless, we have specifically considered the Service time, a mixed distribution of examination, diagnostic, and surgery procedure times. With the help of this strategy, we have attempted to model the Service time of emergency surgical patients. We further discuss the Service time in-depth in section 3.8. After analysing distributions for Interarrival and Service time, we needed to simulate an emergency department (E.D) queuing system based on the priority of patients. The preemptive scheduling approach has many sequence techniques, but the highest response ratio contributes to the priority-based algorithm [20]. It is an approach that considers the ratio of waiting time to burst time for calculating the response ratio. Then this ratio is used to assign the following process to the processor. Waiting time is considered for calculating the response ratio. Ensuring a function with a more significant waiting time will get a higher response ratio, and it will never starve. In 2015, [21] proposed a Varying Response Ratio Priority algorithm. It calculates the priority of each process based on their waiting time and remaining burst time. The algorithm ensures that there will be no endless waiting by waiting time to calculate the priority.

Negi et al. proposed an enhanced Round-Robin algorithm using dynamic time quantum and improved average waiting time [22]. This approach uses a time margin to check whether the process in its second last round could be finished in the same round by extending the time quantum by a small margin. If possible, it will be completed in the same round by providing a time extension for achieving better performance. Vass et al. [23] used the M/M/3 queue model to manage problems like long waiting time, small waiting rooms and insufficient personnel. This approach showed that queuing model could be utilised as an essential tool for decision making. Management of these problems have financial influence, but they also increase service quality while using existing resources. Moreno-Carrillo et al. looked at how the queuing model may be used in the triage process to enhance patient care in the emergency room (ER). The study demonstrated that adopting the model in planning the daily staff requirements in the (ER) triage area can decrease pre-triage time by 65% without increasing additional cost [24]. Modifying the time slice at the beginning of each round is another dynamic version of Standard Round Robin. This method calculates the time slice concerning the residual burst times (RBT) in the successive cycles. The shortest job first

will benefit the processes which are sequenced increasingly based on their burst times (i.e., the process with the highest burst time will be at the tail of the ready queue, and the process with the lowest burst time will be at the head of the row) [25]. The present study used preemptive priority scheduling to provide service to emergency patients according to their priority and severity levels. The reason for using a priority-based preemption technique is to facilitate the decision-makers and clinicians to manage people on the waiting list of their institutions and the patients with low priority so the care could be continued and completed as soon as the server is available [26,27]. This study experienced by applying the preemptive priority scheduling technique, wait time of lowest priority patients, can significantly be reduced if the sufficient number of servers are available according to arrival rate and traffic intensity at each shift runs in the emergency department (ED). This would help in setting the priority rule for eligible patients which demand the multi-skilled resources at the same time [28].

2. THE MODEL

2.1 Description of the Model

The model devised by us has three main objectives in view:

1) To find out a proper distribution of Interarrival time of emergency surgical patients in a public sector hospital of Pakistan.
2) To find out a proper distribution of Service time of emergency surgical patients in a public sector hospital of Pakistan.
3) To simulate the priority queuing system of emergency patients.

An Emergency Department (henceforth referred to as E.D.) has various processes implemented as part of its routine operation, namely registration, examination, surgery procedures and departure. All these processes are accompanied by either FIFO or priority patient queues. The different nature of queues emerges from different risk levels of emergencies. Thus, in (ED), those who come with the highest risk and severely ill are treated first.

2.2. M/M/S Queuing Model of Emergency Department

In the current study, the emergency department (ED) model follows the M/M/s queuing system. M/M/s queueing system is where more than one parallel server serves a single queue. Using Kendall notations, Fig. 1 shows several parallel servers serving a single queue of patients.
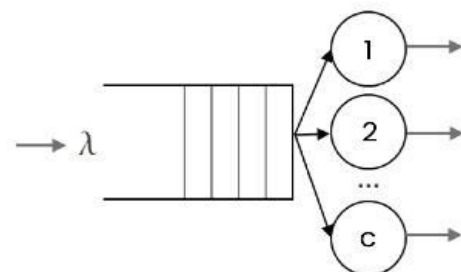

Figure 1: (ED) model follows the M/M/s queuing system

We have used the following variables in our model:

Table 1: Description of Identified variables

variable	Description	Formula
T_p:	represents arrival time of patients	$T_p = T_{p-1} + A_p$.
A_p :	represents an Interarrival time of patients	$A_p = T_p - T_{p-1}$
T_{svc}:	represents Service time of emergency patients	
S:	Number of servers	S_1, S_2
P:	represents the priority of emergency patients	P_1, P_2, P_3

In the present study, the arrival process is a Poisson process. We observed the following three services given to emergency surgical patients: Examination, Diagnostics and Surgical procedures. In the examination (s_1), the patient is examined by the casualty medical officer (CMO) and then the surgeon. Service2 (s_2) refers to the immediate medical tests (diagnostics), and the last service is the surgical procedure (s_3) which an emergency surgical patient receives from the emergency department (ED), shown in Table 4. The description of the identified variables is shown in Table 1.

The distribution of patient arrivals followed a Poisson distribution. The mean value of Poisson parameter λ was 828, with a mean arrival rate of one patient per two minutes. We found Interarrival time exponentially distributed with parameter $1/\lambda = 0.07$ while emergency surgical patients' Service time had gamma distribution, with shape and scale parameters 1.3 and 109. Chi-Square Goodness of fit test was used to identify the emergency surgical patients' distributions of Interarrival and Service time.

2.3. The Study Location

For our case study, we chose The Trauma Centre (the emergency department) of **Abbasi Shaheed Hospital** (a government hospital), located in Paposh Nagar neighbourhood of Nazimabad, Karachi, Sindh, Pakistan. It is the third-largest public sector hospital that comes under the umbrella of the City Government of Karachi. This overcrowded trauma centre operates 24 hours a day and deals with an average of 700 patients daily to give them timely medical care.

There are three shifts in the trauma centre with different staffing levels in all three, shown in Table 2. There are 40 beds altogether in the Trauma Centre, 20 for male and 20 for female, including 5 ICU surgical beds. There are 5 ventilator machines and 2 operation theatres. The ward assistants follow the instructions given by the Clinical Medical Officers (CMOs) or medical residents. The instructions include maintaining the intra-venous (I.V) line levels, administering drips and cannula where applicable, checking temperature, blood pressure and other vital physiological statistics, etc.

2.4. Process Model

Fig. 2 illustrates a new patient's initial examination and the standard clinical response based on the symptoms diagnosed.

Upon entering the hospital, the patient must fulfil the hospital's legal formalities at the registration desk. If a patient is brought in a state of emergency, then the registration formalities are skipped later. A Casualty Medical Officer (CMO) examines the patient. Suppose the patient's symptoms provide evidence of a traumatic injury or illness that cannot be cured or treated in this hospital; the patient is then directed or referred to another hospital. Otherwise, if the patient treatment is available, a senior doctor of the department will examine the patient further. If the patient is deemed non-critical, immediate treatment, if available, is given to the patient and is then discharged timely.

On the contrary, a critical patient, for example, a gunshot victim, is admitted into the hospital and goes under surgical procedure/s if needed. After the successful completion of the surgery, if the patient remains in a critical condition, then he/she is moved to the Intensive Care Unit (ICU). But if the patient is stable enough, then he/she is transferred to the public/private room, depending upon the availability and whichever one the patient can afford. The patients in ICU are also moved to the public/private room once they are stable. Afterwards, depending on the patient's condition, he/she is eventually released from the hospital after getting the doctor's permission.

Table 2: Number of Human Resource in each shift in the Trauma Centre

Shifts	Staffing / Human Resource
Morning	2 casualty medical officers (CMOs), 2 sisters, 1 head nurse, 4 ward boys, 1 senior doctor, 1 surgeon including 3 to 4 house officers, 1 anaesthetist and 2 O.T. technicians. Morning shift works from 0900-1500 hours.
Evening	4 casualty medical officers (CMOs), 2 sisters, 4 ward boys, 1 senior doctor,1 surgeon including 3 to 4 house officers, 1 anaesthetist and 2 O.T. technicians. This shift works from 1500-2200 hours.
Night	2 casualty medical officers (CMOs), 2 sisters, 4 ward boys, 1 senior doctor,1 surgeon including 3 to 4 house officers, 1 anaesthetist and 2 O.T. technicians. This shift works from 2200-0900 hours.

2.5. Emergency Surgery

Emergency surgery is surgery to treat trauma or severe illness. Emergency surgery includes unplanned surgery for patients admitted and unexpected surgery for patients already waiting for an optional surgery (e.g., Obstructive Jaundice). These surgical patients are strictly referred to as "in-patients" since these patients are critical and admit to the hospital for post-operative medical care. We define Emergency surgery as a three colour-casualty system: red, yellow, and green, which can be used to prioritise the patient to be treated, shown in Table 3. Patients use a variety of services based on surgery priorities. Some use three services, and some do not. For example, a patient of priority P_1 having perforation surgery

due to a gunshot may use only two services (s_1,s_3) whereas a patient of priority P_2 may use three services (s_1,s_2,s_3). A description of the services given to the emergency surgical patients in the emergency department is shown in Table 4.
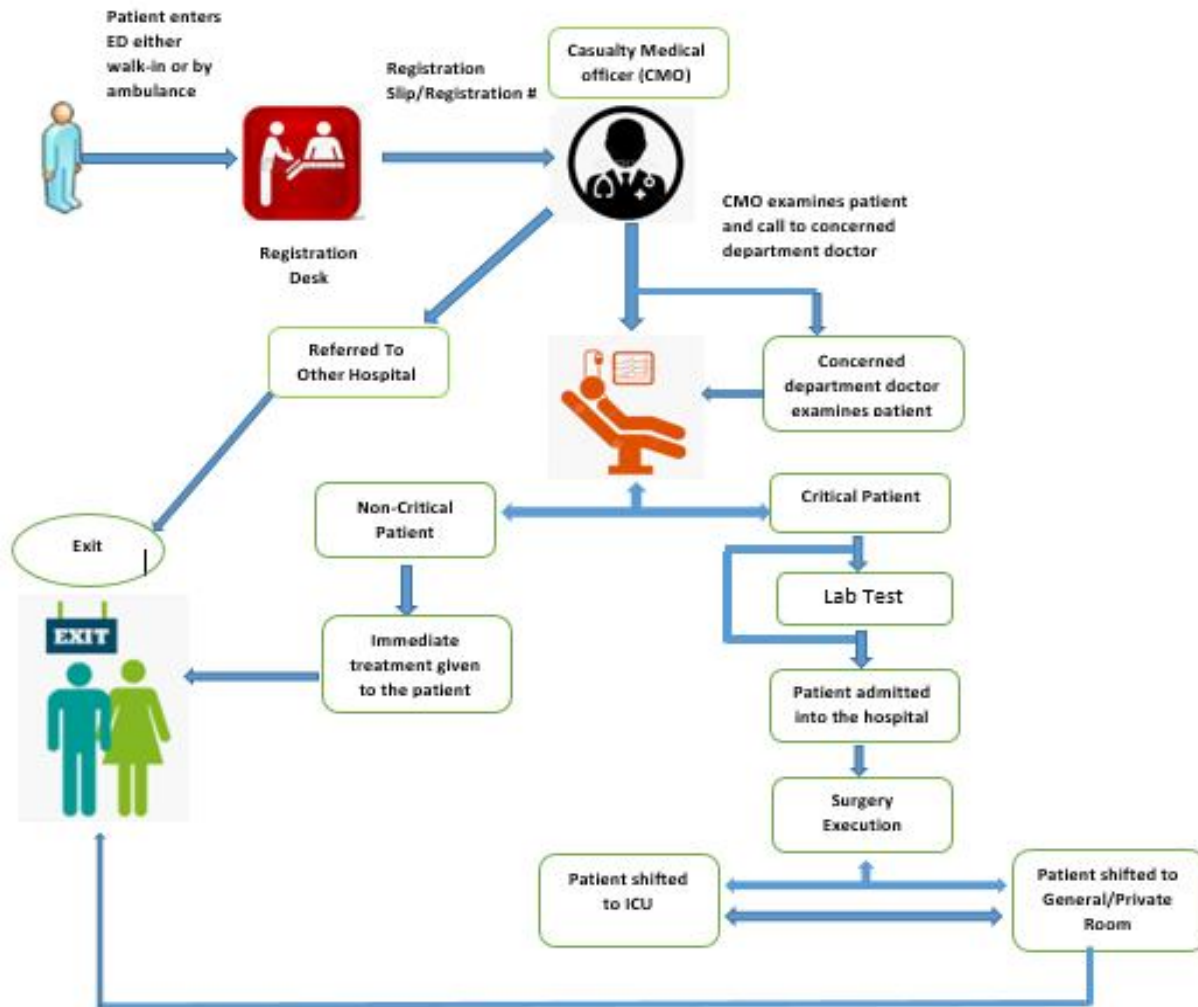

Figure 2: Business process modelling of trauma centre (Abbasi Shaheed Hospital)

Table 3: Priority of emergency surgical patients (colour-casualty system)

Colour Casualty	Priority (P_n)	Priority Description
1	P_1	Represents the most hazardous and critical emergency surgery, for example, perforation surgeries due to tuberculosis, gunshot etc.
2	P_2	Represents the less critical surgeries as compared to red ones. In the yellow category, patients can be in a wait state for some time based on diagnostics or sometimes it is required to check the fitness of the patient to have a surgical procedure. These surgeries include appendectomy, herniotomy, herniorrhaphy, cholecystectomy and other related surgeries of the same kind.
3	P_3	Indicates less severity when contrast with the red and yellow categories of emergency surgery. This kind of surgeries includes incision and drainage (I&D), very commonly, lumpectomy and polypectomy.

Table 4: Description of services

Service Type	Variable used	Service Description
1	S_1	Examination
2	S_2	Diagnostics
3	S_3	Surgical Procedure

3. ANALYSIS

3.1. Horizon of the Analysis

We have based our study on the distribution of Interarrival time and the Service time of emergency surgical patients that arrive 24 hours a day in three shifts (morning, evening, and night). For our observation, we have recorded the arrival of patients for four weeks.

3.2. Data Source

We have collected our subject data for this research at the Abbasi Shaheed Hospital, Karachi, Pakistan. We have recorded our data for one month period during all three shifts per day.

3.3. Chi-Square Goodness of Fit Test

We have utilised the **Chi-square Goodness of Fit Test** as a testing distribution method in the current study. This method determines how well the observed value of a given phenomenon differs significantly from the expected value. This test determines how appropriate the theoretical distribution is in the experimental distribution. We have divided the sample data into multiple intervals and then compared the number of points in these intervals with the expected number of points in each interval.
Following is the equation for Chi-Square Goodness of Fit Test Statistic:

$$X^2 = \sum_{i=1}^{n} \left[\frac{(O-E)^2}{E} \right] \sim X^2_{k-p-1,\alpha} \qquad (1)$$

degrees of freedom
Where
O = Observed value(s)
E = Expected value(s)
k = Number of categories or groups or intervals
p = Number of estimated parameters

3.4. Decision Rule

The decision rule is based on either using the p-value procedure or using the classical technique. In the first procedure, it is determined whether or not the p-value is smaller than α. If the P-value is even smaller than α, reject H_o and accept H_a whereas in the classical approach, it is determined whether or not the measured or calculated test statistic is in the critical region that is H_o is rejected when the calculated or measured value of the Chi-Square statistic is

greater than the tabulated value of the statistic based on degrees of freedom; otherwise, it is accepted.

3.5. Distributions

One way to identify the nature of quantitative data stored in a variable is to study how the data is distributed in the system. We can then get an idea of how the information is distributed in the system with the help of a frequency distribution curve Fig. 3 shows the patient arrivals in the (ED) for 24 hours. On the x-axis, patients' arrival time was stored, and the number of patient counts was on the Y-axis. This curve showed that the patient arrival rate was exceedingly high in all shifts and the maximum number of patients arrived in the second shift.
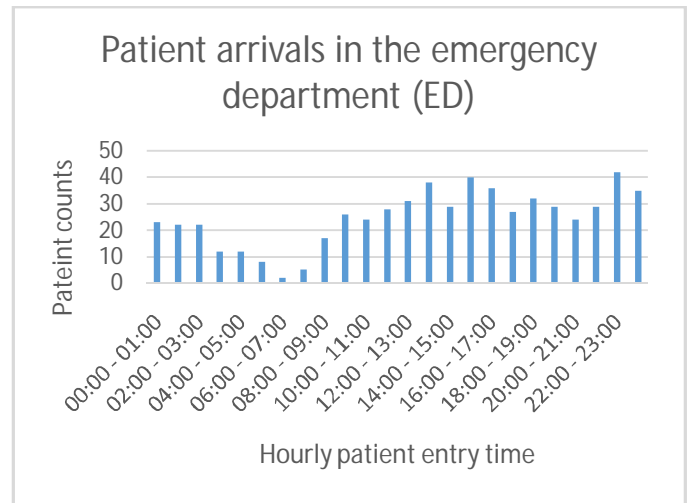

Figure 3: The frequency distribution curve of patient arrivals

Therefore, in our current study, we have modelled the distributions of Interarrival time and Service time of surgical patients in the emergency department (ED). In section 4, exponential and gamma distributions are fitted to the interarrival and Service time of patients.

3.6. Inter-Arrival Time Distribution

Ho: Variable A_p follows an exponential distribution
Ha: Variable A_p does not follow an exponential distribution
The probability density function for the exponential distribution is

$$f(A_p) = \lambda e^{-\lambda * A_p} \quad \lambda > 0 \qquad (2)$$
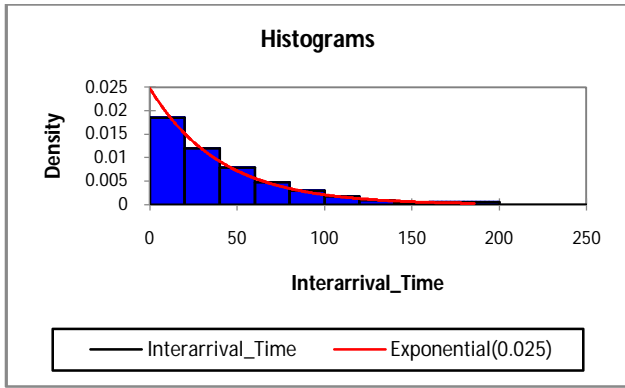
where $\lambda = 40.313$

Figure 4: Histogram of Interarrival time of patients

Table 5: Data Presentation of Interarrival time of patients in minutes

Arrival Time	Arrival Time in minute	Interarrival time	Interarrival time in minute
12:04 AM	4	0:00	0
12:04 AM	4	0:00	0
12:05 AM	5	0:01	1
12:09 AM	9	0:04	4
12:09 AM	9	0:00	0
12:11 AM	11	0:02	2
12:14 AM	14	0:03	3
…	…	…	….
…	…	…	….

Table 5 contains the Interarrival time of patients in minutes. The data collected for 30 days, including three shifts per day. Fig. 4 shows a histogram of Interarrival time of patients in minutes

3.7. Exponential Distribution

The mean value of exponential distribution can be found as $\frac{1}{\lambda}$ that is $\frac{1}{40.313} = 0.0248$.

Therefore, λ is approximately equaled to 0.025.

$$P(a < Ap < b) = \int_a^b 0.025e^{-0.025Ap} \qquad (3)$$

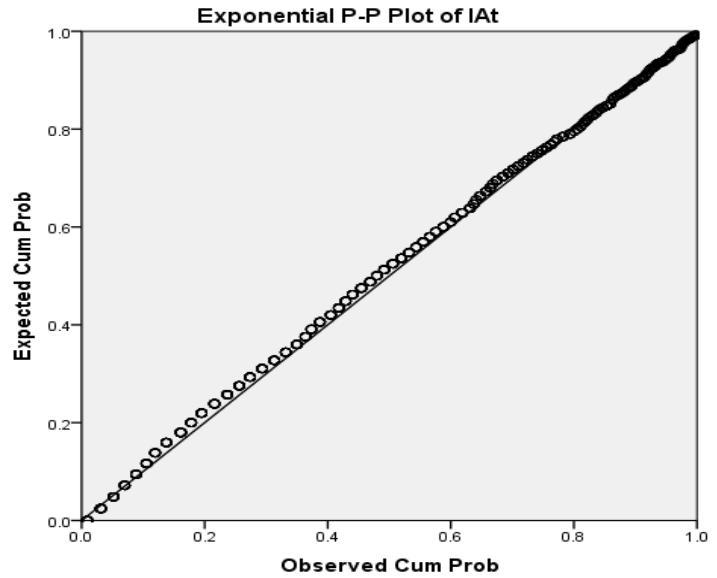$$= e^{-0.025b} - e^{-0.025a}$$

Figure 5: Probability plot of Inter arrival time of patients

In Table 6, we observed the frequency of Interarrival time of patients and compared it with the expected or theoretical frequency, which appeared significantly closer to each other. The last column of Table 6 was calculated using the formula of chi-square goodness of fit test statistic defined in section 3.3.

The test statistic is given as

$$X^2 = \sum_{i=1}^n \left[\frac{(O-E)^2}{E}\right] \sim X^2_{10-1-1, 0.05} \qquad (4)$$

$$X^2 = 10.22 < 15.51$$

Table 6: Chi-square goodness of fit test for the Interarrival time of patients

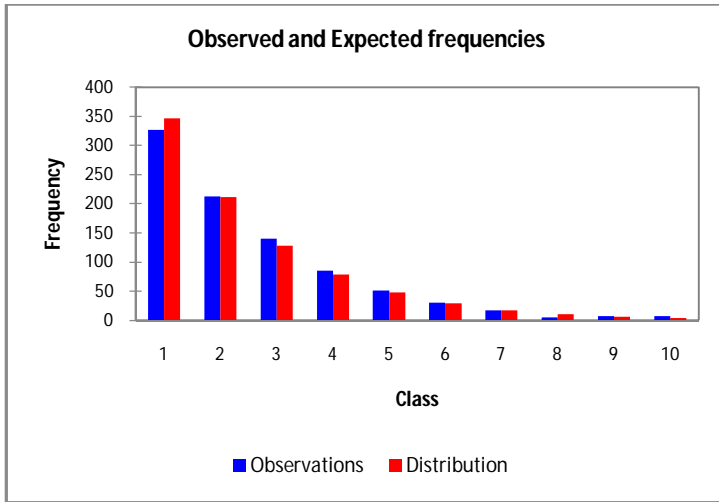| Class	Lower bound	Upper bound	Observed Frequency	Probability	Expected Frequency	Chi-square
1	0	20	329	0.39	350	1.24
2	20	40	213	0.24	212	0.00
3	40	60	140	0.14	129	1.00
4	60	80	85	0.09	78	0.62
5	80	100	52	0.05	47	0.46
6	100	120	31	0.03	29	0.18
7	120	140	17	0.02	17	0.01
8	140	160	6	0.01	11	1.97
9	160	180	8	0.01	6	0.40
10	180	200	8	0.00	4	4.36
Sum			889	1.0	883	10.22784

Figure 6: Comparison of observed and expected frequencies

3.7.1 The Decision Using the Classical Procedure

Conclusion about H_o:
     Fail to reject H_o.
Conclusion about H_a:
     At the 0.05 level of significance, the evidence does not allow us to reject the idea that the Interarrival time of patients in the emergency department (ED) follows an exponential distribution.

3.7.2 The Decision Using the P-Value Procedure

The p-value for the test statistic was found to be 0.215, which was greater than the level of significance,α. Hence there is no way to reject the null hypothesis and to conclude that the Interarrival time distribution is exponential.

3.8. Service Time Distribution

Ho: Service time (St) of emergency surgical patients follows a gamma distribution.
Ha: Service time (St) does not follow gamma distribution; the Service time of emergency surgical patients does not have a gamma distribution.
The probability density function for the gamma distribution is given as

$$f(St, \alpha, \beta) = \frac{\beta^\alpha St^{\alpha-1} e^{-\beta St}}{\Gamma_\alpha} \qquad (5)$$

where α is a shape parameter, and β is known as the scale parameter of the gamma distribution.
Table 7 represents the collection of Service time of 186 emergency surgical patients in minutes, where P1 –P186 represents a patient number. Each cell in the last column of the table below, against the patient number, shows either the summation of two or three services shown in Table 4. All of the services (s_1, s_2, s_3) followed exponential distribution. Hence the summation of two or more exponentially distributed services proved to follow a gamma distribution.

Table 7: Data representation of Service time of emergency surgical patients.

Patient #	S_1	S_2	S_3	Service time
p_1	10	0	26	36
p_2	6	143	41	190
p_3	6	90	110	206
p_4	10	0	29	39
p_5	9	80	110	199
p_6	7	0	59	66
….	….	….	….	….
p_n	….	….	….	….
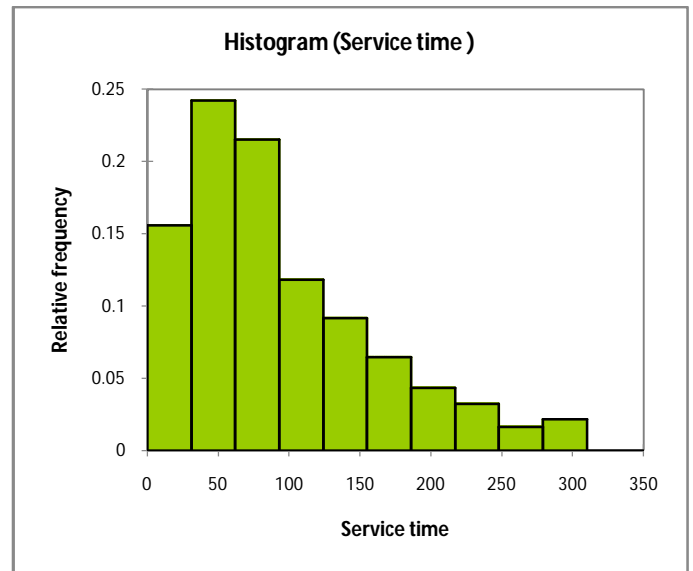
Figure 7:Histogram of Service time of emergency surgical patients

3.9. Gamma distribution

$$\alpha = \left(\frac{\mu}{S}\right)^2 \qquad (6)$$

$$\beta = \frac{S^2}{\mu} \qquad (7)$$

Calculation of shape and scale parameters:
     mean = μ = 91.575
Standard Deviation = S = 68.539
Variance = S^2 = 4697.608
Substituting the values in equation number 6 and 7, we get
$\alpha = \left(\frac{91.575}{68.539}\right)^2, \beta = \frac{4697.608}{91.575}$

$\alpha = 1.785, \qquad \beta = 51.297$

Figure 8: Probability plot of Service time of emergency surgical patients

hence, accept that emergency surgical patients' Service time follows a gamma distribution.

Figure 9: Distribution of Service time of emergency surgical patients
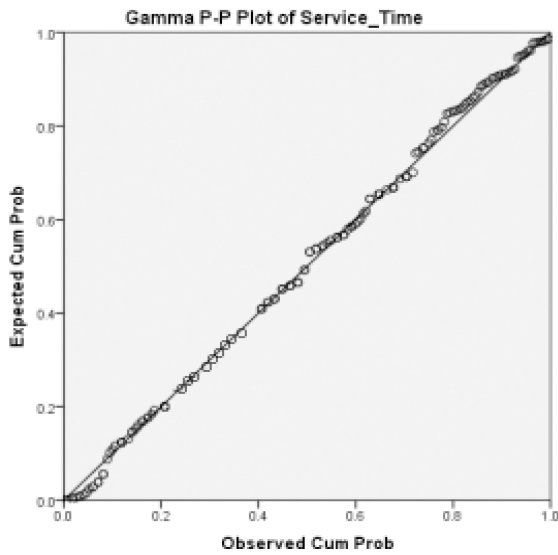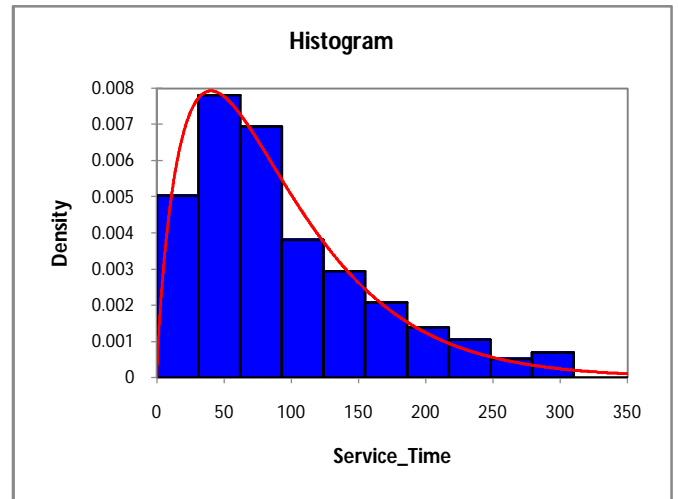
3.9.1. The Decision Using the Classical Procedure

Conclusion about H_o:
      Fail to reject H_o.
Conclusion about H_a:
      At the 0.05 level of significance, the evidence allows us to reject the alternative hypothesis and,

Table 8: Chi-square goodness of fit test for Service time of emergency surgical patients

Class	Lower bound	Upper bound	Observed Frequency	Probability	Expected Frequency	Chi-square
1	0.000	31.000	29	0.16878471	31.386	0.181
2	31.000	62.000	45	0.24010325	44.656	0.003
3	62.000	93.000	40	0.19847249	36.916	0.258
4	93.000	124.000	22	0.14196703	26.408	0.736
5	124.000	155.000	17	0.09475649	17.627	0.022
6	155.000	186.000	12	0.06071754	11.295	0.044
7	186.000	217.000	8	0.03787233	7.046	0.129
8	217.000	248.000	6	0.02317524	4.312	0.661
9	248.000	279.000	3	0.01398047	2.601	0.061
10	279.000	310.000	4	0.0083408	1.552	3.862
Total			186	0.98817034	184	5.957

3.9.2. The Decision Using the P-Value Procedure

The p-value for the test statistic was found to be 0.744, which was more significant than the level of significance,

α. Hence there is no way to reject the null hypothesis. Therefore, we have concluded that the Service time follows a gamma distribution.
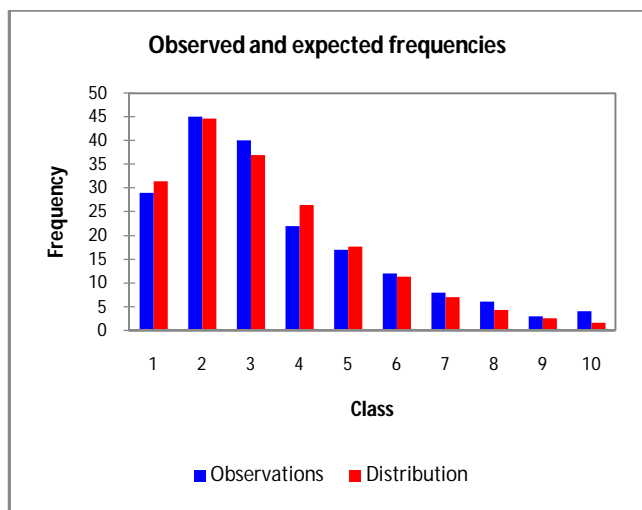
Figure 10: Comparison of observed and expected frequencies

4. SIMULATION MODEL

This study has built a simulation model for simulating a priority queuing system of emergency patients for only one service (s1). We observed the arrival of patients to E.D. for an hour with an arrival rate of 30 patients per hour. That is one patient per two minutes. The service rate that we observed was 7.6 minutes.

4.1. Analysis of A Service (S1) Given to the Patients In (ED) Based On Priority

- Patients arrive at random times from {0,1, …,4} minutes.
- Service time s vary from {3, 4...,12} minutes.
- Consider the system for 15 patients.
- Two servers/ night shift.

Table 9: Arrival, Service time and priority of emergency patients.

S.no	Patient Number	Arrival Time	Inter Arrival	Priority	Service time (s1)
1	p1	0	0	3	10
2	p2	4	4	2	6
3	p3	5	1	2	6
4	p4	9	4	3	10
5	p5	9	0	3	9
6	p6	11	2	2	7
7	p7	14	3	3	12
8	p8	16	2	3	10
9	p9	17	1	2	7
10	p10	17	0	3	8
11	p11	20	3	1	3
12	p12	22	2	2	6
13	p13	24	2	1	5
14	p14	25	1	3	7
15	p15	25	0	3	8
Total		218.0	25.0		114

4.2. Simulation Runs for 15 Patients

Simulation runs as follows:

Step 1:

- For Patient p, generate an Interarrival time A_p. Add it to the previous arrival time T_{p-1} to get the arrival time of patient p as

$$T_p = T_{p-1} + A_p. \quad (8)$$

Step 2:

- If Server 1 (CMO1) is idle, Patient p starts service with the server S_1 (CMO1) at the current time T_{now}.
- Server1 (S_1)'s completion time $T_{fin,}$ is given by

  $$T_{fin,S_1} = T_{now} + T_{SVC,S_1} \quad (9)$$

  where T_{SVC,S_1} the Service time is generated from S_1's Service time distribution. Patient p's waiting time is

$$T_{wait} = 0, \text{ where } T_{wait} = T_{sys} - T_{SVC} (10)$$

- Patient p's time in the system T_{sys}, is given by

  $$T_{sys} = T_{fin,S_1} - T_p. \quad (11)$$

- If S_1 is busy and S_2 is idle, Patient p starts with S_2 (CMO2). The rest is the same.
- If the new Patient p, arrived at the ready queue has a higher priority than the currently being examined Patient, the server (CMO) is preempted, which means the examination service (S_1) of the current patient is stopped, and the incoming new patient with higher priority gets the server (CMO) for the examination service (s_1).

Step 3:

- If S_1 and S_2 are both busy, then calculate the time at which the first one becomes available as follows:

  $$T_{beg} = min (T_{fin,S_1}, T_{fin,S_2}). \quad (12)$$

- Patient p begins service at T_{beg}. When service for Patient p begins, set T_{now} as

  $$T_{now} = T_{beg}. \quad (13)$$

- Compute T_{fin,S_1} or T_{fin,S_2} as in step 2.
- Response Time,

  $$T_{res} = T_{now} - T_p \quad (14)$$

- Patient p's time in the system is given by

  $$T_{sys} = T_{fin,S_1} - T_p \text{ or } T_{sys} = T_{fin,S_2} - T_p$$

4.3. Preemptive Priority Queuing System in (ED)

Section 4.3 illustrates how patients were scheduled in the emergency department (ED) based on the severity and priority levels to be treated first. Fig. 12(a) and 12(b) have used a Gantt chart to show order and how the service was provided to patients based on their priority, to get examined by the doctor at the earliest. For service s_1, two servers were available at night shift, S_1 and S_2 respectively. A separate Gantt chart has been created for both servers so that an individual chart can specify the utilisation of the respective server. In addition to which the starting time and ending time of each patient's service can also be achieved with the help of this Gantt chart

which illustrated the characteristics of preemptive priority scheduling. Consider the utilisation of S_1 which started its utilisation by providing the service to the first patient who arrived in E.D. at time 0 minute with priority = 3 and remained to provide the service for five minutes while ′s service duration was 10 minutes. It was just because S_1 was preempted when the third patient p_3 arrived in the ready queue with priority number 2, which was higher than the priority of the first patient, so preemption took place and S_1 was switched to p_3. When preemption took place, S_2 had already started examining p_2 arrived at time 4 minutes, which had the same priority as of p_3. Both $p_2 and p_3$ had same priority, therefore S_2 did not switch toward p_3 as patients having the same priority are served based on a first come, first serve (FCFS). Again S_1 preempted while serving p_4 for 7 minutes because

of p_{11}, whose priority was 1, which was highest among all the preferences shown in Table 3. The service of p_4 was interrupted twice because of higher priority patients in the ready queue. Hence those patients whose service was interrupted had experienced a delay in getting the service finished on time. When the service of p_4 was interrupted a second time by S_2 , it was finally completed by the same server whereas S_1 initiated to serve p_4 very first time. It indicates that if the service of any patient is preempted, it can be continued and completed by any available server in the system. Consider the utilisation of S_2 which was less than the utilisation of S_1 as shown in the Gantt chart and also in the simulation table because S_2 had been consumed for a smaller number of minutes as compared to S_1
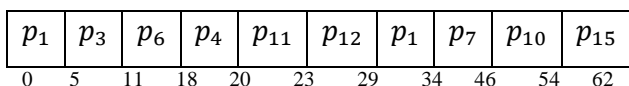
p_2	p_4	p_9	p_{13}	p_4	p_5	p_8	p_{14}
4	10	17	24	29	30	39	49	56

Figure 11 (b): Preemptive priority queue (S_2)

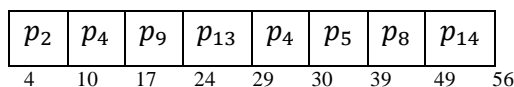| p_1	p_3	p_6	p_4	p_{11}	p_{12}	p_1	p_7	p_{10}	p_{15}
0	5	11	18	20	23	29	34	46	54	62

Figure 11 (a): Preemptive Priority queue (S_1)
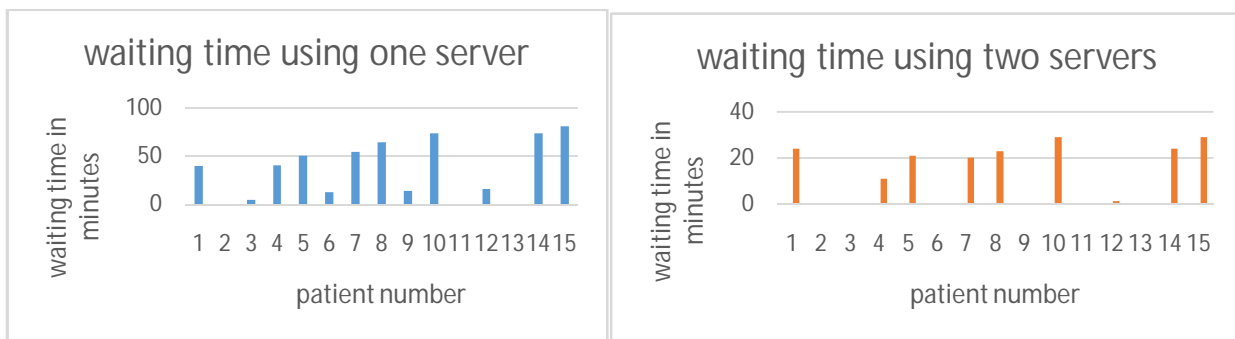

Figure 12 (a): Wating time of patients using one server **Figure 12 (b):** Waiting time of patients using two servers

4.4. Simulation of Priority Queue of Patients In Emergency Department (ED)

Table 10: Simulation Table

p#	At	Priority	Service time	Completion Time	p spent time in the system	Wait Time	Response Time
p_1	0	3	10	34	34	24	0
p_2	4	2	6	10	6	0	0
p_3	5	2	6	11	6	0	0
p_4	9	3	10	30	21	11	9
p_5	9	3	9	39	30	21	21
p_6	11	2	7	18	7	0	0
p_7	14	3	12	46	32	20	20
p_8	16	3	10	49	33	23	23
p_9	17	2	7	24	7	0	0
p_{10}	17	3	8	54	37	29	29
p_{11}	20	1	3	23	3	0	0
p_{12}	22	2	6	29	7	1	1
p_{13}	24	1	5	29	5	0	0
p_{14}	25	3	7	56	31	24	24
p_{15}	25	3	8	62	37	29	29
Total	**218**		**114**		**296**	**182**	**156**

4.5. Server Utilisation

Table 11: Server Utilisation

S.no	patient #	Arrival	Priority	when S_1 available	when S_2 available	server chosen	Service start S_1	Service end S_1	Service start S_2	Service end S_2	Active S_1	Active S_2
1	p1	0	3	1	1	S_1	0	5			5	0
2	p2	4	2	0	1	S_2			4	10	0	6
3	p3	5	2	1	0	S_1	5	11			6	0
4	p4	9	3	0	1	S_2			10	17	0	7
5	p6	11	2	1	0	S_1	11	18			7	0
6	p9	17	2	0	1	S_2			17	24	0	7
7	p4	9	3	1	0	S_1	18	20			2	0
8	p11	20	1	1	0	S_1	20	23			3	0
9	p12	22	2	1	0	S_1	23	29			6	0
10	p13	24	1	0	1	S_2			24	29	0	5
11	p1	0	3	1	1	S_1	29	34			5	0
12	p4	9	3	0	1	S_2			29	30	0	1
13	p5	9	3	0	1	S_2			30	39	0	9
14	p7	14	3	1	0	S_1	34	46			12	0
15	p8	16	3	0	1	S_2			39	49	0	10
16	p10	17	3	1	0	S_1	46	54			8	0
17	p14	25	3	0	1	S_2			49	56	0	7
18	p15	25	3	1	0	S_1	54	62			8	0
										$\sum T_{fin}$	62	52

Table 12: Server utilisation results

Server	Utilisation Formula	Value (%)
S_1	$U_{S_1} = \dfrac{\sum T_{fin,S_1}}{\sum T_{svc}}$	54.38
S_2	$U_{S_2} = \dfrac{\sum T_{fin,S_2}}{\sum T_{svc}}$	45.61

4.6. Simulation Results

Table 13: Simulation results for each priority of emergency patients

Priority	T_{sys} (minutes)	T_{wait} (minutes)	T_{res} (minutes)
1	4	0	0
2	6.6	0.2	0.2
3	31.87	22.62	19.37

Table 14: Simulation results for all priorities of emergency patients

Variable	Description	Formula	Result (minutes)
T_{sys}	The average time a patient spends in the system for s_1	$T_{fin} - T_p$ OR $T_{wait} + \bar{s}$	19.73 12.13 + 7.6 = 19.73
T_{wait}	Average wait time	$T_{sys} - T_{svc}$	12.13
T_{res}	Response time	$T_{now} - T_p$	10.4
P_{wait}	The probability that a patient has to wait	$\dfrac{\text{Number of patients who wait}}{\text{Number of patients}}$	0.6
$\bar{s}$	Service time	$\dfrac{\sum \text{Service time}}{\text{Number of patients}}$	7.6
$\bar{\lambda}$	The average time between arrivals	$\dfrac{\sum A_p}{\text{Number of } T_p - 1}$	1.785
$\bar{W}_{waited}$	Average waiting time of those who wait	$\dfrac{\sum T_{wait}}{\text{Number of patients that wait}}$	20.22

Table 15: Comparison of results obtained using simulation with one and two servers

Variable	Value (with one server)	Value (with two servers)
T_{sys}	42.58	19.53
T_{wait}	35.26	12.13
T_{res}	32.06	10.4
P_{wait}	0.8	0.6
$\bar{s}$	7.6	7.6
$\bar{\lambda}$	1.785	1.785
$\bar{W}_{waited}$	44.08	20.22

5. CONCLUSION

In this study we have analysed the Interarrival and the Service time of the emergency surgical patients, in one of the busiest public sector hospitals of Pakistan. Improving the efficiency of services required for emergency patients are always on higher priority for the emergency department (ED) of any hospital. Since the patients in (ED) always arrive unexpected with uncertain conditions, the need of medical treatment in timely fashion is always critical. One of the best ways to improve the service quality in hospitals' (ED) is to use a simulation model. With the help of queuing model of simulation, arrival, and service pattern can be modelled well. The present study has selected the chi-square goodness of fit test to identify emergency surgical patients' interarrival and Service time distribution. Table 6 and 8 show the test statistic results of the "Goodness of fit test" revealing that the Interarrival time follows an exponential distribution while Service time of emergency surgical patients follows a gamma distribution. This Service time is composed of two or three kinds of services that are exponentially distributed. Hence, the summation of two or three exponentially distributed services proves to be gamma distribution. This analysis of Service time reveals that emergency surgery patients are served with two or

more kind of services. The patients waiting for emergency surgery and the patients of other medical conditions may face a situation due to some severity and priority levels to be treated first. In our simulation model the emergency surgical patients utilise only one service that is examination service (s_1), at night shift. (s_1) is the firstservice that any emergency patient receives in (ED) while diagnostics (s_2) is secondservice if needed, the patient is then recommended for a surgical procedure (s_3). Following the simulation results shown in Table 13 and 14, it is observed that the arrival rate is remarkably high while the number of servers is considerably low due to which average wait time is 35.26 keeping server utilisation at 99.99 % as shown in Table 15. In our simulation model, we have chosen 15 patients and among them only two patients had highest priority (priority = 1). In general, the arrival rate was 30 patients per hour. Had there been a larger number of high priority patients arrived within an hour, than the mid and lower priority patients had to wait even longer to be examined. Table 14 shows simulation results for each priority level, indicating that the highest priority level patients treated without waiting for a single minute in the queue. The lowest priority level patients suffered

a longer waiting time, because the maximum number of the patients in the queue had lowest priority. So, in future, if a large number of highest priority patients come with the same ratio of lowest priority patients, than the number of referrals would increase, and the overall efficiency of the (ED) would decrease. Table 15 shows the comparison between the simulation results obtained using one and two servers, it displays significant reduction in waiting time for patients, by simply employing two servers, hence urges the need for more servers to manage critical treatment and emergency care for both in and outpatients during each shift in the emergency department (ED). The arrival process follows a Poisson process with one patient every two minutes. This situation demands more servers especially during the night shift since traffic accidents and gunshot cases often arrive at night.

6. FUTURE WORK

In future, this study would help producing simulation of services needed to provide to patients in the emergency department (ED), gathering accurate results and statistics so well-informed decisions could be made to improve the overall efficiency of the emergency department (ED).

REFERENCES

[1] J.C. Moskop, D.P. Sklar, J.M. Geiderman, R.M. Schears and K. J. Bookman. **Emergency Department Crowding, Part 1—Concept, Causes, and Moral Consequences**, *Annals of Emergency Medicine*, vol. 53, no. 5, pp. 605-611, May 2009.

[2] N.R. Hoot and D. Aronsky. **Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions**, *Annals of Emergency Medicine,* vol. 52, no. 2, pp. 126-136, July 2007.

[3] P. Hall and J. Park. **Nonparametric inference about service time distribution from indirect measurements**. *Journal of the Royal Statistical Society: Series B*, vol. 66, no. 4, pp. 861–875, October 2004.

[4] N.H. Bingham and S.M. Pitts. **Non-parametric Estimation for the M/G/∞ Queue**. *Annals of the Institute of Statistical Mathematics*, vol. 51, pp. 71-97, March 1999.

[5] S. Schweer and C. Wichelhaus. **Nonparametric estimation of the service time distribution in the discrete-time GI/G/∞ queue with partial information**, *Stochastic Processes and Their Applications*, vol. 125, no. 1, pp. 233–253, January 2015.

[6] K. I. Park and Park. *Fundamentals of Probability and Stochastic Processes with Applications to Communications*, New Jersey, USA: Springer International Publishing, 2018.

[7] P. Billingsley, *Probability and Measure*, 3RD ed. New York, USA: Wiley, 1995.

[8] Yong-Hong Kuo, Janny M. Y. Leung, Kelvin K. F. Tsoi, Helen M. Meng and Colin A. Graham. **Embracing Big Data for Simulation Modelling of Emergency Department Processes and Activities**, in *proc. 2015 IEEE International Congress on Big Data IEEE*, New York, 2015, pp. 313–316.

[9] Y. H. Kuo, J. M. Y. Leung and C. A. Graham**. Simulation with data scarcity: developing a simulation model of a hospital emergency department**, in *proc: Proceedings of the 2012 Winter Simulation Conference (WSC), IEEE,* Berlin, 2012, pp. 1–12.

[10] M.M. Günal and M. Pidd. **Discrete event simulation for performance modelling in health care: a review of the literature,** *Journal of Simulation,* vol. 4, pp. 42-51, March 2010.

[11] L. Luo, Y. Zhang, J. Yang and H. Guo. **Modeling patient arrivals in a radiology department**, in *proc. 2017 International Conference on Service Systems and Service Management, IEEE,* Dalian, 2017, pp. 1–5.

[12] L. V. Green, S. Savin and B. Wang. **Managing Patient Service in a Diagnostic Medical Facility**, *Operations Research.* vol. 54, no. 1, pp. 11–25, February 2006.

[13] P. Bhattacharjee and P. K. Ray**. Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections**, *Computers & Industrial Engineering*, vol. 78, pp. 299–312, December 2014.

[14] N. T.J. Bailey. **A Study of Queues and Appointment Systems in Hospital Out-Patient Departments, with Special Reference to Waiting-Times**, *Journal of the Royal Statistical Society: Series B* (*Methodological),* vol. 14, pp. 185–199, July 1952.

[15] S.L. Albin, J. Barrett, D. Ito and J.E. Mueller. **A queueing network analysis of a health center**, *Queueing Syst, vol.* 7, pp. 51–61, March 1990.

[16] M. E. Zonderland and R. J. Boucherie. **Queuing Networks in Health Care Systems**, In *Handbook of Healthcare System Scheduling*, vol. 168, Boston: Springer, 2012, pp. 201-243.

[17] A.S. Oladimeji and O.J. Ibidoja.**The Distribution of Service Time of Patients**, *Journal of Reliability and Statistical studies*, vol. 13, no. 1, pp. 61–72, October 2020.

[18] I.P. Lade, S.A. Chowriwar and P.B. Sawaitul. **Simulation of Queuing Analysis in hospital**, *International Journal of Mechanical Engineering and Robotic Research*, vol. 2, no. 3, pp. 122–128, July 2013.

[19] B.K. Abbas. **Simulation Models of Emergency Department in Hospital**, *Journal of Engineering and Development.* vol. 18, no. 2, pp. 172–179, March 2014.

[20] H.M. Deitel. *An Introduction to Operating Systems*, New Jersey, USA: Addison-Wesley Publishing Company Inc, 1984.

[21] P. Singh, A. Pandey and A. Mekonnen. **Varying Response Ratio Priority: A Preemptive CPU Scheduling Algorithm (VRRP)**, *Journal of Computer and Communications,* vol. 3, no. 4, pp. 40–51, March 2015.

[22] S. Negi. **An Improved Round Robin Approach using Dynamic Time Quantum for Improving Average Waiting Time**, *International Journal of Computer Applications*. vol. 69, no. 14, pp. 12–16, May 2013.

[23] H. Vass and Z.K. Szabo. **Application of Queuing Model to Patient Flow in Emergency Department. Case Study**, *Procedia Economics and Finance*, vol.32, pp. 479–487, December 2015.

[24] A.Moreno-Carrillo, L.M.Á. Arenas, J.A. Fonseca, C.A. Caicedo, S.V. Tovar, O.M. Muñoz-Velandia**. Application of queuing theory to optimize the Triage process in a tertiary emergency care ("ER") department**, *Journal of emergencies, trauma, and shock*, vol. 12, no. 4, pp. 268–273, Oct-Dec 2019.

[25] A. Pandey, P. Singh, N. H. Gebreegziabher and A. Kemal.**Chronically Evaluated Highest Instantaneous Priority Next: A Novel Algorithm for Processor Scheduling**, *Journal of Computer and Communications,* vol. 4, no. 4, pp.146–159, April 2016.

[26] J. Déry, A. Ruiz, F. Routhier, M.-P. Gagnon, A. Côté, D. A.-Kadi, V. Bélanger, S. Deslauriers, M.-E. Lamontagne.**Patient prioritization tools and their effectiveness in non-emergency healthcare services: a systematic review protocol**, *Systematic reviews*, vol. 8, pp. 1-7, March 2019.

[27] J. Dery, A. Ruiz, F. Routhier, V. Bélanger, A. Côté, D. Ait-Kadi, M.-P. Gagnon, S. Deslauriers, A.T.L. Pecora, E. Redondo and A.-S. Allaire.**A systematic review of patient prioritization tools in non-emergency healthcare services**, *Systematic reviews*, vol.9, pp. 1– 4, October 2020.

[28] N. Wongwai, S. Malaikrisanachalee, J. Prajongmoon, **Performance of Heuristic Priority Rule for Multiskilled Resource-Constrained Scheduling**. *International Journal of Emerging Trends in Engineering Research*, vol. 9, pp. 633-642, June 2021.