



OPERATION RESEARCH: QUEUING ANALYSIS IN HOSPITAL

Hospital Administration

Dr GC Gulati Commander, 1 Technical Training Wing.

Dr Nitin Balram Ahuja* Directing Staff, Officers Training College. *Corresponding Author

Dr Priyanka Singh Pulmonologist, Command Hospital.

ABSTRACT

Queuing theory is an operational research technique that helps administrators in planning demand supply mis-match in service industries. Application of Queuing theory in hospitals will go a long way in planning facilities in a resource constrained environment as well as enhancing clientele satisfaction. The spread used in the article has demonstrated how waiting time can be calculated, and how service facilities can be placed. The Queue formation is a phenomenon which occurs at most instances in our daily lives, especially when demand for a particular service exceeds the capacity of the system that has to provide the service. Using the queuing theory, this mis-match can be minimized and optimal utilization of resources can be ensured.

KEYWORDS

Queue, waiting time, operational research, demand supply mis-match

INTRODUCTION

The arena of operations research provides a scientific approach to managerial decision making. In a contemporary, competitive ever-changing corporate world, a manager needs quantitative techniques and a factual way of solving problems related to optimal allocation of resources. Medical Administrators, are skeptic in dealing with the operation research as it emphasizes on mathematics. Hence effort is to the application part of the operation research with minimal emphasis on arithmetic. The health care industry all over the world is overwhelmed with delays and queue formation, we are used to long queues in the general OPDs, specialist OPDs, pharmacy service and hospital bed allocations or availability. There are several consequences to waiting i.e. delayed care, poor patient satisfaction, greater financial implications etc. Queuing theory deals with delays caused by the discrepancy between demand for a service and the capacity to meet the demand. During 1903 Mr. A.K. Erlang, a Swedish engineer has started theoretical analysis of waiting line problem in telephone calls. In 1927, Mr. Millins developed the theory further and then by Mr. Thornton D Fry. But Mr. D.G.Kendall has given a systematic and mathematical approach to waiting line problem in 1951. Queuing theory can be applied to a variety of conditions where it is not possible to precisely predict the arrival rate (or time) of patient and service rate (or time) of hospital service facility or facilities.

The Structure Of A Queuing System

The major apparatuses of any waiting-line /queuing system are shown in Each of these components is discussed below:

- Calling population (input source)
- Queuing process
- Queue discipline
- Service process (or mechanism)

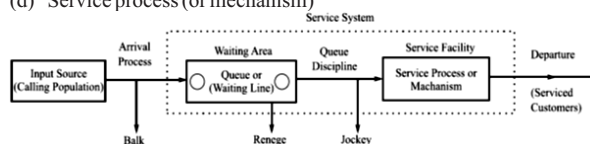


Figure 1: Queuing System

Potential patients who arrive to the queuing system is referred as calling population, also known as patient (input) source. The manner in which patients arrive at the service facility, individually, or in batches, at scheduled or unscheduled time is called the arrival process. The patient's entry into the queuing system depends upon the queue conditions.

Patients, from a queue, are selected for service according to certain rules known as queue discipline. A service facility may be without server (self-service), or may include one or more servers operating either in a series (as a team) or in parallel (multiple service channels). The rate (constant or random) at which service is rendered is known as the service process. After the service is rendered, the patient leaves the

system. If the server is idle at the time of the patient's arrival, then the patient is served immediately, otherwise the patient is asked to join a queue or waiting line, which may have single, multiple or even. The manner in which Patients arrive at the service facility, individually, or in batches, at scheduled or unscheduled time is called the arrival process. If the server is idle at the time of the patient's arrival, then the patient is served immediately, otherwise the patient is asked to join a queue or waiting line, which may have single, multiple channels.

Calling Population Characteristics. The calling population need not be homogeneous and may consist of sub population as patients arriving in an OPD of a hospital can be a walk-in patient, a patient with the prior appointment or as an emergency case. The behaviour of the patient arrival pattern can we generally classified as follows. However, if a patient customer, on arriving at the service system waits in the queue until served and does not switch between waiting lines, he is called a patient customer

- Balking** Patients do not join the queue either by seeing the number of patients already in service system or by estimating the excessive waiting time for the desired service.
- Reneging** Patients, after joining the queue, wait for some time in the queue but leave before being served on account of certain reasons.
- Jockeying** Patients move from one queue to another hoping to receive service more quickly

Pattern of arrivals at the system Patients may arrive in batches or individually. These patients may arrive at a service facility (OPD counter) either on scheduled time (by prior information) or on unscheduled time (without information). The arrival process (or pattern) of patients to the hospital service system is usually approximated to **Poisson distribution** which is a discrete probability distribution, describes the arrival rate variability, i.e., number of random arrivals at a service facility in a fixed period of time. The Poisson distribution has single parameter, called (the Greek lowercase letter λ), which is the mean or expected number of events per unit. The variance of a Poisson distribution is also equal to, and the standard deviation is equal square root of λ . The number of events, X , of the Poisson random variable ranges from 0 to infinity (∞).³

$$P(X) = \frac{e^{-\lambda} \lambda^X}{X!}$$

$P(X)$ = the probability of X events in an area of opportunity

λ = expected number of events

e = mathematical constant approximated by 2.71828

X = number of events.

$X!$ = is the factorial of X

Considering an empirical example of Poisson distribution of patient reporting to Accident & Emergency Department of a hospital where in patients arrive at a hospital emergency center at the rate of 3 patients

per hour and there is need to ascertain the probability of two patients arrivals during the next 30 minutes?

λ = The arrival rate during 30 minutes = $(3/60)*30 = 1.5$.
 n = The random variable = 2 occurrences

A	B	C
Hosp Arrival Mean per hour	λ	3
2 patient in half hour	λ	1.5
No of Occurance	Exact Probability	Cumm Probabil
0	0.22	0.22
1	0.33	0.56
2	0.25	0.81
3	0.13	0.93
4	0.05	0.98
5	0.01	1.00
6	0.00	1.00
7	0.00	1.00
8	0.00	1.00
9	0.00	1.00
10	0.00	1.00

Figure 2: Poisson Distribution

Using Poisson Distribution function in Excel The probability of exactly 2 arrivals in the next 30 minutes is 25 %. The graph is skew to the right

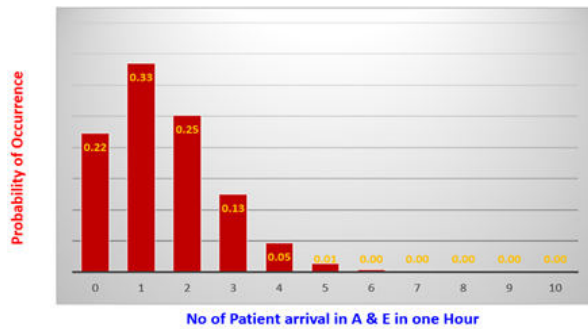


Figure 3: Arrival Probability Factor

Queuing Process The queuing process refers to the number of queues – single, multiple or priority queues and their lengths. The type of queue depends on the layout of service mechanism and the length (or size) of a queue depends upon operational situations such as physical space, legal restrictions, and attitude of the patients.

Queue Discipline The queue discipline is the order (or manner) in which patients from the queue are selected for service. There are a number of ways in which patients in the queue are served. If customers are served in the order of their arrival, then this is known as the first-come, first served (FCFS) service discipline. Pre-emptive priority (or Emergency) Under this rule, a critical patients may be allowed to enter into the service immediately after entering into the system, even if a patient with lesser priority are already in service.

Arrangement of Service Facilities The capacity of the service facility is measured in terms of customers who can be served simultaneously and / or effectively. The service facilities (or servers), commonly known as service channels, may be in series or in parallel

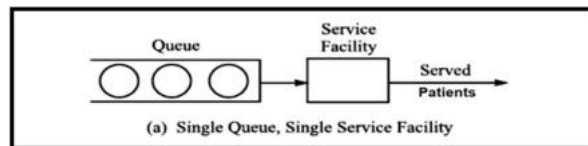


Figure 4: Single Queue service facility

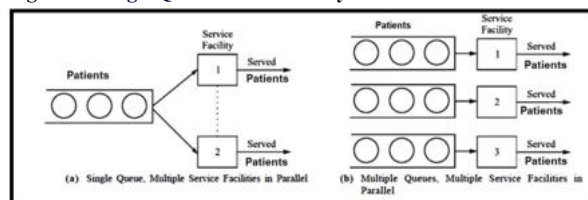


Figure 5: Multiple Facility

Performance Measures Of A Queuing System The performance measures (operating characteristics) for the evaluation of the

performance of an existing queuing system, and for designing a new system in terms of the level of service a customer receives as well as the proper utilization of the service facilities are listed as follows:

(a) Average (or expected) time spent by the patient in the queue and system

Wq : Average time an arriving customer has to wait in a queue before being served,

Ws : Average time an arriving patients spends in the system, including waiting and service.

(b) Average (expected) number of patients in the queue and system

Lq : Average number of patients waiting for service in the queue (queue length)

Ls : Average number of patients in the system (either waiting for services in the queue or being served).

(c) Value of time both for patients and servers

Pw : Probability that an arriving patients has to wait before being served (also called *blocking probability*). $\rho = \lambda$

μ : Percentage of time a server is busy serving patients, i.e., the system utilization.

Pn : Probability of n patients waiting for service in the queuing system.

Pd : Probability that an arriving patients is not allowed to enter in the queuing i.e., system capacity is full.

Therefore, the queuing formula for single channel and single phase is as under

	In Line	In Service	In the System
Expected Time	$Wq = \lambda / \mu (\mu - \lambda)$	$1 / \mu$	$W = 1 / \mu - \lambda$
Expected Number	$Lq = \lambda^2 / \mu (\mu - \lambda)$	$\rho = \lambda / \mu$	$L = \lambda / \mu - \lambda$

Assuming that the patient arrival at an Accident & Emergency of a large teaching hospital is in according to a Poisson process with a mean rate of 5 cases per hour. The Service time is exponentially distributed with an average of 8 minutes per patient. The hospital management is interested in determining the performance measures for this service system. Populating the data into excel sheet as follow

A	B	C	D
Arrival Rate λ		5	
Service Rate μ		7.5	
Utilisation Factor $\rho = \lambda / \mu$		0.6667	
Time	1hr	60	
Expected Time in the Line Wq		0.2667	16
Expected Time in the Service Ws		0.1333	8
Expected Time in the System $1 / \mu - \lambda$		0.4	24
Expected No in the Line Lq		1.3333	1
Expected No in the Service Ls		0.6667	1
Expected Number in the System $\lambda / \mu - \lambda$		2	

Figure 6: Performance Measures

Figure 7: Programming the cells of the Excel sheet

Hence it is evident that if average patient arrival is five casualty per hour in a large teaching hospital, with the service time being eight minute per patient in the accident and emergency department of a hospital the expected time for the patient in the line is about 16 minutes, the expected time in the service is 8 minutes and the total waiting time in the system is about 24 minutes.

Arrival Rate λ	Service Rate μ	Utilisation Factor ρ	Length in the Queue
1	8	0.13	0.02
2	8	0.25	0.08
3	8	0.38	0.23
4	8	0.50	0.50
5	8	0.63	1.04
6	8	0.75	2.25
7	8	0.88	6.13

Figure 8: Expected Time

If the arrival rate is gradually increasing and the service rate continues to remain the same the utilization factor rho (ρ) which is λ/μ increases assume and at the one point length of the queue $Lq = \lambda^2 / \mu(\mu - \lambda)$ move toward infinity as per the graph of which is depicted as below this could be rho value 0.75.

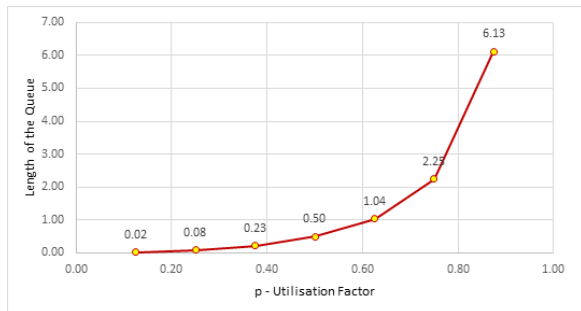


Figure 9: Matrix

CONCLUSION:

Operations research (OR) is a discipline that deals with the development and application of advanced analytical methods to improve decision-making. It is considered as a subfield of mathematical sciences and is widely used in management science. Queuing models can help forecast responses to queries about patient flow, such as how long will the average patient have to wait to avail a service or how long will it take for an average visit in a health care institution. In medical setup, queuing models are generally based on three factors ie patient arrival rate, server rate (service time for clinical exam, treatment, etc.) and the number of servers (clinical and nonclinical staff) available.

REFERENCES

1. P. Rama Murthy, Operation Research (2nd Edition) 2007, NEW AGE INTERNATIONAL (P) LIMITED, PUBLISHERS, 4835/24, Ansari Road, Daryaganj, New Delhi - 110002 New Age International (P) Ltd., Publishers, 450-472
2. JK Sharma, Operation Research -Theory and Application. 2017, 6th Edition, Trinity Press, 113, GOLDEN HOUSE, DARYAGANJ, NEW DELHI - 110002, INDIA
3. David M. Levine David F. Stephan, Timothy C. Krehbiel Mark L. Berenson (2008), Statistics for Managers: Using Microsoft Excel, Fifth Edition, Pearson Custom Publishing