Emerging Markets Queries in Finance and Business

Application of Queuing Model to Patient Flow in Emergency Department. Case Study

Hajnal Vass[a], Zsuzsanna K. Szabo[b]

[a]Mures County Emergency Hospital-Emergency department, Gh. Marinescu street no.50, Targu Mures 540136
[b]"Petru Maior" University of Targu Mures, N. Iorga street no 1, Targu Mures 540088

Abstract

Emergency Department (ED) overcrowding represents a common characteristic that may affect the quality and access to health care. Analyzing the ED presentations during the last few years we have observed a constant increasing of presentation numbers. For each ED it is a challenge to decrease the patients' waiting time, to provide timely care and to improve the patient's satisfaction. Long waiting times is the most important complaint in patient**'s** satisfaction surveys. We have evaluated 2195 questionnaires for a period of three years (2010-2012). The general satisfaction rate is 84, 63% and the most frequent complaints are about the waiting time which is too long, the waiting room which is small and the personnel which is insufficient. To manage these situations in a proper way we proposed, for our analysis, to use queuing models which can provide reasonably accurate evaluations of our system's performance. The data used in the case study includes detailed information's over the period January 1 - December 31, 2012, when totally 51.458 patients were registered. The results of this study can help us to understand the magnitude of the broader problem, the relationship between resources and waiting times, and to provide a method for understanding and monitoring performance, to find solutions for a better understanding and for alleviating the daily crisis.

Keywords: Emergency Department; Queuing Theory, Waiting time;

1. Introduction

Overcrowding in Emergency Departments (ED) is a problem worldwide and affects the ability to provide emergency medical care within a reasonable period of time. The number of patients who present in emergency departments is growing and the department ability to assist patients with acute complains is constant. Delays in the ED may cause dramatic outcomes for patients. ED's performance in terms of patients flow and of the available resources can be studied using the Queuing Theory. Emergency medical system can be regarded as a network of queues and different types of servers where patients arrive, wait for a service, get a result and then go home or they are admitted to a hospital unit. The waiting threads are effective tools to support management decisions and the hospitals could benefit from the results of these researches.

Research papers show that the Queuing Theory can be used efficiently in health care. The Queuing Theory was first analyzed by Agner Krarup Erlang, a Danish engineer, mathematician, researcher, in 1913 in the context of telephone traffic. The uses of this theory in health care are relatively recent. Most articles were published after 1990, and that due to the advancement in computational power and software availability.

Researchers have shown that the queuing theory can be useful in health care. Lakshmi C, Sivakumar Appa Iyer reviews the contributions and applications of the queuing theory in the field of health care management problems. Samuel Fomundam and Jaffrey Herrmann (Fomundam & Herrmann, 2007) summarize a range of queuing theory results in the following areas: waiting time and utilization analysis, system design and appointment system. McClain (McClain, 1976) reviews research on models for evaluating the impact of bed assignment policies on utilization, waiting times and the probability of turning away patients. Seshaiah and Thiagaraj (Seshaiah & Thiagaraj, 2011) using a queuing model, they obtained a relationship between the percentage of reneged patients and the reneging parameter in addition to finding the wait time distribution. Albin et al. (Albin, Barett, D.Ito, & Muller, 1990) use the QNA software, which calculates server utilization, the mean and variance of the number of customers at each node, the time that patients are in a multi-node network, the mean and the variance of waiting time at each node. De Vericourt and Jennings (Vaericourt & Jennings, 2011) present a queuing model to determine efficient nurse staffing policies. Mayhew and Smith (Mayhew & Smith, 2008) illustrates how flow though an accident an Emergency department can be represented as a queuing process, how the outputs of this model can be used to visualize and interpret the 4-hours Government target in a simple way. Worthington (D.J.Worthington, 1991) suggests that increasing service capacity has little effect on queue length. McQuarrie (McQuarrie, 1983) find out that is possible, when utilization is high, to minimize waiting time by giving priority to those who require shorter service time. Hoot and Aronsky (N.R.Hoot & Aronsky, 2008) conducted a systematic review with the main objective to describe the scientific literature on Emergency Department crowding from the perspective of causes, effects, and solutions.

In this paper in order to plan the capacity of Emergency Department situated in Mures County, Romania, to manage in optimal way the patients flow, queuing models will be used. We proposed, for our analysis, to use queuing models to provide reasonably accurate evaluations of the system's performance.

2. Model of patient flow

It is common for health care managers to project workload for physical infrastructure and manpower planning. They must consider five typical measures when evaluating the existing or proposed service systems. These measures are:

o average number of patients waiting (in queue or in the system);
o average time the patients wait (in queue or in the system);
o capacity utilization;

o costs of a given level of capacity;

o probability that an arriving patient will have to wait for service.

The system utilization measure reflects the extent to which the servers are busy rather than idle. On the surface, it might seem that health care managers would seek 100% system utilization. Under normal circumstances, 100% utilization may not be realistic; a health care manager should try to achieve a system that minimizes the sum of waiting costs and capacity costs. In queue modeling, the health care manager also must ensure that the average arrival and the service rates are stable, indicating that the system is in a steady state, a fundamental assumption.

The main queuing model characteristics are (Yasara, 2009):

o the population source;

o the number of servers;

o the arrival patterns and the service patterns;

o the queue discipline.

The population source can be infinite or finite. In an infinite source situation, patient arrivals are unrestricted, and can exceed the system's capacity at any time.

Number of servers - the capacity of the queuing systems is determined by the capacity of each server and the number of servers being used.

Arrival patterns – the waiting lines occur because highly variable arrivals and service patterns cause the systems to be temporarily overloaded. The hospital ED is very good examples of random arrival patterns causing such variability. The arrival pattern is different at different times of the day.

Service patterns - because of the varying nature of the illnesses and the patients' conditions, the time required for treatment varies from patient to patient.

Queue discipline refers to the order in which customers are processed. The assumption that service is provided on a first-come, first-served basis is the most commonly encountered rule. The ED does not serve on this basis, patients do not all represent the same risk, level of triage; those with the highest risk, the most seriously ill, are treated first.

The queue system is usually described in shorten form by using some characteristics. These characteristics can be represented by Kendall's notation which was initially a three factor notation A/B/C. Later D, E and F were also included in the model to make it A/B/C/D/E/F. The notation of the queuing model is presented in Table 1:

Table 1. Notation of the queuing model

Symbol	Explanation
A	The arrival time distribution
B	The service time distribution
C	The number of servers (agent available)
D	The system's capacity, the number of customers in the system
E	The calling population
F	The queue discipline

There are some special notations that have been developed for various probability distributions describing the arrivals and departures. Some examples are: M - Arrival or departure distribution that is a Poisson process, E - Erlang distribution, G - General distribution, GI - General independent distribution.

For the application of the queuing models to any situation we should describe first the input and the output process. In our ED the input process is the patient's arrival and the output process is considered the patient's discharge or the admission in the hospital unit.

In order to build the flowchart of the patients' accessing and departing the ED we will use the queuing theory to determine the minimal number of servers (providers) needed. Queue models generally deal with customer arrivals at a service facility.

We will consider an M/M/n queuing model because it will help us to estimate the number of providers needed. Arrivals occur according to a Poisson process and the service duration has an exponential distribution. Poisson distribution is a discrete distribution that shows the probability of arrivals in a given time period, where the mean and the variance of the Poisson distribution are the same. Using this M/M/n model we assume that:

$$\frac{\lambda}{n\mu} < 1, \tag{1}$$

where we denote by:

λ - arrival rate;
μ - service rate;
n - number of server (provider);
ρ - system utilization;
$1 / \mu$ - service time;
P_0 - probability of 0 units in system;
P_k =-probability of k units in system.

To optimize the process we are looking for the probability P_k the probability than an entering patient must queue for treatment which means that all physicians are busy. In order to calculate these probabilities we will use relations:

$$P_k = P_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{k!} \qquad\qquad k < n \tag{2}$$

$$P_k = P_0 \left(\frac{\lambda}{\mu}\right)^k \frac{1}{n! n^{k-n}} \qquad\qquad k \geq n \tag{3}$$

Based on the property that the overall sum of probabilities must verify:

$$\sum_{k=0}^{\infty} P_k = 1 \tag{4}$$

we will calculate P_0 and the probabilities P_i for any i. We mention that until the number of arrivals *k* is less than the number of servers (providers) *n,* we do not have queue (waiting line). But if the number of arrivals exceeded the number of servers (providers) then the process must be optimizes to reduce the waiting time.

3. Queue Model for Emergency Department. Case study

In Romania the emergency access of the patients is realized through ED which belongs to the hospitals. The Emergency Department is designed to continue the treatment begun in the street, at the accident site, or to begin the treatment if the patient has arrived with a non-medical vehicle, to diagnose the patient in cooperation with the specialists of the hospital and to observe the patient until the hospital admission or until his release, in case the patient's condition does not require hospital treatment.

For the case study we considered the ED which functions on the ground floor of the Mures County Emergency Hospital. This ED is one of the most modern units in Romania, and provides a high quality care for patients who require evaluation and treatment for a period of time between 0 and 24 hours. Having an area of approximately 2,200 square meters, it operates in a well delimited space, equipped with the latest medical equipment. The emergency department consists of several major areas: triage, resuscitation room, immediate care unit, space for minor emergencies, room for minor surgeries, pediatric emergencies, Computer Tomography, critical observer.

The data used in the case study includes detailed information's over the January 1- December 31, 2012. During the analyzed period 55.12% of the patients were brought by ambulance services, out of which 8,68% needed priority treatment at their arrival and 44. 88% of the treated patients were walk -ins. 34% of the treated patients in the ED required hospitalization.

The working process in the ED: the first patient-personnel interaction occurs in the triage area, where a triage nurse evaluates the patient, determines the seriousness of the patient's health condition and assigns a corresponding triage level or code which ranges from 1 to 5 as can be seen in table 2. Patients with level 1 proceed directly to the resuscitation room and those with level 2 to the immediate care unit. Patients with levels 3, 4, 5 move to the waiting area unless a bed is immediately available.

Table 2. Triage levels (triage codes) in ED

Triage level / code	Waiting time to first medical consult
Level I / Red code	0 minute
Level II / Yellow code	Maxim 10 minute
Level III / Green code	Maxim 30 minute
Level IV / Blue code	Maxim 60 minute
Level V / White code	Maxim 120 minute

The workflow of the ED will be analyzed as an absolute queuing process, the cases with red code are treated by priority, in which the patients arrive, wait, are evaluated and treated and then they are discharged or are transferred in the hospital units.

Building on the patient flow we will use the queuing theory to estimate the necessary number of ED resources (human resources), to estimate the average waiting time. The results are important for the management of the ED department in order to make optimal decisions, to organize in optimal way the workflow.

In the analyzed period a total of 51.458 patients were registered, thus the daily average (24 hours) can be considered as 150 patients. We created a data base with the patients' name, address, arrivals time, diagnosis and time spent in ED.

We denoted by x_i the annual average number of arrivals of the patients in the time interval $[t_{i-1}, t_i]$, where $t_0=0$ means the 0 hour, the beginning of the day. If we denote $P_k^{d,i}$ the number of patients arrived

on day d in the time interval $[t_{i-1}, t_i]$, then the average number of the patient's arrival/day we obtain from:
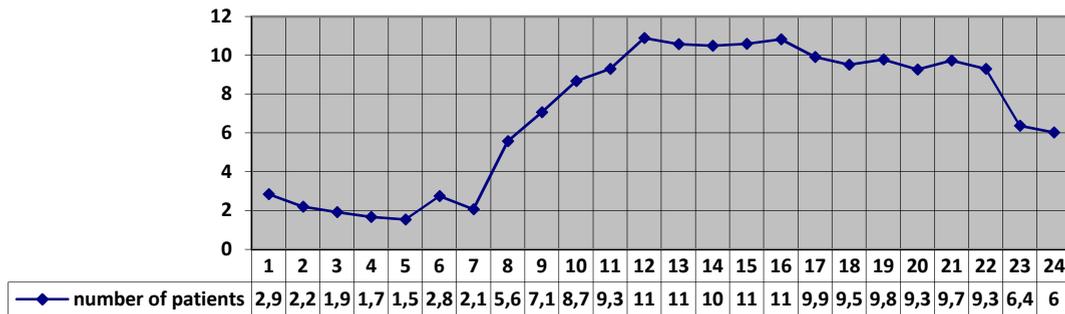
$$x_i^d = \frac{\sum_{M=1}^{12} \sum_{d=1}^{\{29,30,31\}} P_k^{d,i}}{12} \tag{5}$$

where M it means the months and the sum after d refers to the number of days of each month. Thus, the average number of the patients arrived in the interval $[t_{i-1}, t_i]$ we will get from the relation:

$$x_i = \frac{x_i^d}{24} \tag{6}$$

The calculated values of x_i, the considered time intervals $[t_{i-1}, t_i]$ can be seen on figure 1. Arrivals of the patients are a random process, as figure 1 shown.

Figure 1 – Average number of arrivals in 2012

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
number of patients	2,9	2,2	1,9	1,7	1,5	2,8	2,1	5,6	7,1	8,7	9,3	11	11	10	11	11	9,9	9,5	9,8	9,3	9,7	9,3	6,4	6

Source: Own calculations, based on our database. Data refer to 2012

Thus, the average number of arrivals denote by lambda will be:

$$\lambda = \frac{\sum x_i}{24} \tag{7}$$

We obtained $\lambda = 7$, thus we will consider in the following that the patients arrive at a rate of 7/hour and stay in a single queue. Using the created database, using same rationality as above, the calculated annual average service rate is: $\mu = 3$, thus each physician needs on average 20 minutes to deal white a patient. It is known that the system is in steady state if the relation is fulfilled:

$$\frac{\lambda}{n\mu} < 1$$

where n presents the number of human resources, the number of physicians in the ED. Thus, using the inequality can be estimated the minimum number of physicians in the ED.

In the case when physician needs on average of 20 minutes to deal with a patient, the number of arrivals per hour λ will determine the minimum number of physicians n, as can be seen in table 3.

Table 3. Minimum number of physician

λ	7	8	9	10	11	12
n calculated	3	3	4	4	4	5

Source: Own calculations

We used M/M/3 queuing model to estimate different specifications of the queue, different characteristics of the ED.
In our case study we have: $\lambda=7$; $\mu=3$; $n=3$; $\alpha=7/3$; and $SE=\lambda/n\mu=7/9$

In the case of M/M/3 model to calculate the probability that no patients is in the ED, we use the condition, that de overall sum of probabilities must be 1. We can write:

$$P_0 + P_1 + (P_2 + P_3 + \cdots) = 1 \tag{8}$$

We substitute the probabilities, using n=3 and we get:

$$P_0 + P_0 \frac{\lambda}{\mu} + P_0 \left(\frac{\lambda}{\mu}\right)^2 \frac{1}{2!} + P_0 \left(\left(\frac{\lambda}{\mu}\right)^3 \frac{1}{3!3°} + \left(\frac{\lambda}{\mu}\right)^4 \frac{1}{3!3} \cdots \right) = 1 \tag{9}$$

$$P_0 + P_0 \frac{\lambda}{\mu} + P_0 \left(\left(\frac{\lambda}{\mu}\right)^2 \frac{1}{2!} + P_0 \left(\frac{\lambda}{\mu}\right)^3 \frac{1}{3!} \left[\frac{1}{3°} + \left(\frac{\lambda}{\mu}\right)\frac{1}{3} + \left(\frac{\lambda}{\mu}\right)^2 \frac{1}{3^2} \cdots \right] = 1 \tag{10}$$

The geometric series is convergent and introducing the sum of the series we have:

$$P_0 + P_0 \frac{\lambda}{\mu} + P_0 \left(\frac{\lambda}{\mu}\right)^2 \frac{1}{2!} + P_0 + \left(\frac{\lambda}{\mu}\right)^3 \frac{1}{3!} \frac{3\mu}{3\mu-\lambda} = 1 \tag{11}$$

The probability that no patient is in the ED is:

$$P_0 = \frac{1}{1 + \frac{\lambda}{\mu} + \frac{1}{2} \left(\frac{\lambda}{\mu}\right)^2 \frac{3\mu}{3\mu-\lambda}} \tag{12}$$

In the case of ED, based on one year data, the probability that no patient in the Ed, no queues is:
$$P_0 = 0.064$$

In M/M/3 queue an entering patient must queue for service exactly when 3 or more patients are already in the system. Using Erlang's C formula, we get in our case:

$$C\left(3, \frac{\lambda}{\mu}\right) = \sum_{k=3}^{\infty} P_k = 1 - \sum_{k=0}^{2} P_k \tag{13}$$

$$C\left(3, \frac{\lambda}{\mu}\right) = P_0 \frac{1}{3!} \left(\frac{\lambda}{\mu}\right)^3 \frac{3\mu}{3\mu-\lambda} \tag{14}$$

Introducing in relation (14) the data we get:

$$C\left(3,\frac{\lambda}{\mu}\right) = 0.611.$$

The length of the queue in the case of M/M/3 is:

$$L_q = P_0 \frac{1}{3!}\left(\frac{\lambda}{\mu}\right)^3 \sum_{k=1}^{\infty} k\left(\frac{\lambda}{3\mu}\right)^k \tag{15}$$

Calculating the sum of the series we obtain:

$$L_q = P_0 \frac{1}{3!}\left(\frac{\lambda}{\mu}\right)^3 \frac{3\lambda\mu}{(3\mu-\lambda)^2} \tag{16}$$

And using relation (14) we have:

$$L_q = \frac{\frac{\lambda}{3\mu}}{1-\frac{\lambda}{3\mu}}\ C\left(3,\frac{\lambda}{\mu}\right) \tag{17}$$

At the ED the steady state mean number of patients in the queue is L_q=2.13, thus for a new arrival exists a high probability to stay in queue.

Now we are interested to know the waiting time in queue. Using the Little's Law, the mean waiting time in the queue can be obtained from:

$$W_q = \frac{L_q}{\lambda} \tag{18}$$

Using the relation (17) we have:

$$W_q = \frac{1}{3\mu-\lambda}C\left(3,\frac{\lambda}{\mu}\right) \tag{19}$$

At the ED the mean waiting time is W_q=0.30 hour, it means 18 minutes. The average treatment time per patient, denoted by W_t is $W_t = 1/\mu$, thus W_t=0.33 hour (19.8 minutes).

Thus the total waiting time in system can be obtained from: $W=W_q + W_t$. Introducing data we get W=0.63 hour (37.8 minutes).

The overall number of patients in the ED is on average L=Wλ. In the ED based on the data the overall number of patients will be: L =4.41.

4. Conclusions:

In this case study we use M/M/3 queue model to characterize the patient flow in the Emergency Department situated in Mures County, Romania.

We created a data base using 51.458 registered cards, the total number of patients assisted and treated in ED during 2012. The study illustrates how data analysis and queuing models can be used in decision making to find optimal solutions. In the case study we considered that the number of human resources n is the same with the number of beds. In further studies we intend to extend the research for the case when the waiting

time is not limited only to the number of physicians. Practice shows that the majority of the cases the physicians can deal white more than 1 patient in 20 minute. In the ED is very important to determine optimal allocation of the resources, number of physician and the number of beds according to the demands, in order to have optimal waiting time for the patients. The queuing model it is a useful instrument for capacity planning. The planning activity can be realized using computer simulation. Figure 1 shows that a dynamic allocation of the medical staff is needed.

However, as it has been demonstrated over many years, models can be invaluable in providing decision support in complex environments as the ED. This study supports ED managers, the model can be used in the management of the ED as a useful instrument for decision making. These problems have financial impact but in a same time increase the quality of the services. Given the financial constraints that exist, it is very important to find ways to improve performance with the existing resources.

References

Albin, S., Barett, J., D.Ito, & Muller, j. (1990). A queuing network analysis of a health center. *Queuing systems 7* , 51-61.

D.J.Worthington. (1991). Hospital waiting list management models. *Journal of Operational Research Society 42* , 833-843.

Fomundam, S., & Herrmann, J. (2007). A survey of Queuing Theory Applications in healthcare. *ISR Technical Report 24* .

Mayhew, L., & Smith, D. (2008). Using queuing theory to analyse the government's 4-h completion time target in accident and emergency departments. *Health Care management Science 11* , 11-21.

McClain, j. (1976). Bed planning using queuing theory models of hospital occupancy: a sensitivity analysis. *Inquiry 13* , 167-176.

McQuarrie, D. (1983). Hospital utilisation levels. The application of queuing theory to a controversial medical economic prblem. *Minesota Medicine 66* , 679-686.

N.R.Hoot, & Aronsky, D. (2008). Systematic Review of Emergency Department Crowding: Causes, Effects, and Solutions. *Annals of Emergency Medicine 52* , 126-136.

Seshaiah, C., & Thiagaraj, H. (2011). A queuing network congestion model in hospitals. *European Journal of Scientific Research 63* , 419-427.

Vaericourt, f. D., & Jennings, O. (2011). Nurse staffing in medical units: a queuing perspective. *Operation Researche 59* , 1320-1331.

Yasara, O. (2009). Queuing Models and Capacity Planning. In O. Yasara, *Queuing Methods in Health care management* (pp. 348-356). San Francisco: Jossey-Bass.