International Journal of Research in Medical Sciences
Shanmugam R. Int J Res Med Sci. 2014 Aug;2(3):1076-1084
www.msjonline.org
pISSN 2320-6071 | eISSN 2320-6012

DOI: 10.5455/2320-6012.ijrms20140882

Research Article

How do queuing concepts and tools help to efficiently manage hospitals when the patients are impatient? A demonstration

Ramalingam Shanmugam*

Department of Health Administration, Texas State University, San Marcos, TX 78666, USA

Received: 30 June 2014
Accepted: 13 July 2014

***Correspondence:**
Dr. Ramalingam Shanmugam,
E-mail: rs25@txstate.edu

© 2014 Shanmugam R. This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ABSTRACT

Background: Due to severe pain, patients are impatient in several wings sporadically and more frequently in emergency wing of the hospitals. To efficiently administer in such environment and the hospital management seeks helpful strategies. The queuing concepts and related methodologies can help as this article has demonstrated by an analysis and interpretation of real data from a hospital in Malta.
Methods: The queuing concepts are probabilistic and statistical ideas based approach. They require configuration of the rate and pattern of arriving patients, the rate and pattern of the service, the number of channels serving, the capacity of the waiting room, and the criterion for selecting patients for service etc. New ideas are presented in this article to manage in various scenarios of real life emergency operations. The pertinent queuing concepts and tools are made easier for the readers to comprehend and practice in their own situations in which they notice that the patients are impatient in their waiting.
Results: Using the new ideas and formulas of this article, the data in the emergency wing of a hospital in Malta (a largest island of an archipelago situated in the center of the Mediterranean with a total population of a million) are analyzed and interpreted. The results clearly explain why there were a prolonged waiting times at the emergency department creating public dissatisfaction and patients were leaving without waiting to be seen. The total time spent by non-urgent patients with nurse and casualty officer is more in the second shift and lesser and lesser in the third and fourth shifts. The interactive time with a nurse by patient is statistically same in all three types: life-threatening, non-life threatening but urgent, and non-urgent. Very strikingly, the patients in all three groups wait longer to be seen by the nurse in shift three and lesser time in shifts two or four.
Conclusion: In 21st century with flourishing globalized medical tourism, a standardized approach to minimize efficiently the waiting time in emergency and other wings of the hospitals in developing as much as in developed nations is a necessity as this auricle has pointed out. The impediments and the remedies for an efficient standardization are overdue.

Keywords: Busy time, System's memory, Service discipline, Waiting time, Probability distribution

INTRODUCTION

Queue is everywhere in almost all segments of life and it becomes a source of frustrations and monetary loses. The longer queue length turns away even the loyal customers in a business operation. Consequently, the business loses profits. The management has no choice but to quickly resolve the stumbling issue of having a longer queue. In this process of resolving, the management is attracted to a scientific and prudent approach called queuing theory. Alternatively, the queuing theory is called waiting line concepts.

What is waiting time? The waiting time, W of a customer in the system is the sum of waiting time, W_q in the queue and his/her service time, $1/\mu$ where μ is the number of customers received service per unit time. More research articles in refereed journals and educational books have been written on queuing theory. Each approach and its results in queuing theory are based on a model.

What is model? The model is an abstraction of the reality. Using a new model, this chapter illustrates an interesting aspect of virtual waiting time [see http://virtualqueueingwordpress.com and Kampllikar (2005)] which is the perceived time people feel that they wait for service in hospitals. Its impacts are that newly arriving patient might decide not to enter the hospital and the hospital staff might speed up their service to build up a good image. This waiting time is highly influenced for patients due to their situation. A literature review of the topic is first done first and then a case is made for a new model. How important is the waiting time topic. Unlike places in restaurants, airport checking, bank counters, highway entrances, grocery counters etc. where customers wait in queue, the patients waiting in hospitals or clinics for healthcare services exhibit more visible impatience either due to intolerable pain or other reasons. The patients' impatience often receives sympathetic consideration by the queue managers in the hospitals or clinics. Consequently, the medical team gets alerted to quicken their service. This alert level provides advantages to the patients with higher impatience but results in inconveniences to other patient's in the inflow or outflow of the hospital system. Based on a literature search on this topic of patients' impatience, this article compiles, discusses the state of the art on queuing concepts in generality and then presents a few innovative strategies to help improve the hospital's service efficiency since it is vital to successful healthcare business operations.

WHAT ARE QUEUING CONCEPTS AND THEIR TOOLS?

Since birth till the death, everyone in his/her life is an integral part of healthcare system. The healthcare is perhaps the most determinant of both quality and longevity of life. All developed or developing nations have to deal with issues and their solutions of healthcare reforms. Hospitals are meant to shelter patients to receive treatment to recover from an illness to partial/complete healthy status. See Buhaug (2002), Preater (2002), Propper et al. (2002), Sternberg (2006), and Kennedy et al. (2004) for details on issues connected to waiting time in hospitals. Evans (2007) mentions a list of duties for the patients. The patient flow is indicative of efficiency of hospital operation. See Jacobs (2001) for methods to increase hospital efficiency. A poor flow formulates queue and consequently a time delay to receive service occurs. A synchronization of patient's admission, treatment and discharge in a hospital is a necessity to reduce the waiting time for a patient.

What factors are connected to patient's service? Obviously, the scheduling of physicians, nurses, diagnosing labs, insurance payments and pharmacists should match the patients' arrival patterns and medical needs. See Tucker et al. (1998) and Vericourt et al. (2011) for details about the importance of queuing theory to plan a hospital's needs. The waiting time is a mix of productive and non-productive types. For an example, the waiting time to recover from a surgery is a necessity and productive type. Another example is the non-necessary waiting time for the patient's guardian to arrive and approve an emergency treatment and it is non-productive type. See Matthias et al. (2011), Green (2006), Wellstood et al. (2005) and de Bruin (2007) for details on the consequences of long queue in emergency wards and van Klei et al. (2002), Dunnill et al. (2004) in outpatient wards. In an ideal hospital system, the service to a patient ought to be instantaneous. Could the length of stay in a hospital be near zero time? It is doubtful because of limited resources and a necessity to monitor the patient. Hospitals are mandated to operate within local, state and federal regulations. Most importantly, the hospitals have to function within human ethical limitations.

Consider, for an example, the flow of patients to an emergency room (Table 1). The patient's waiting and service times depend on the time of arrival and/or the seriousness of the patient's condition. More details are discussed in the data analysis section of this chapter. However, the patient's service, whether in an emergency or other room, is coordinated through other support units such as radiology, surgery, pharmacy, diagnostic labs, patient records and transportation service. In queue terminology, these services constitute multiple servers. What is queuing theory? The queuing theory dates back to an engineering researcher Erlang in 1908.

Table 1: Waiting time for test results in an England's hospital.

Diagnostic test in	>6 weeks	>13 weeks	>26 weeks
Audiology	0.72	0.57	0.44
Colonoscopy	0.35	0.18	0.13
Neurophysiology	0.44	0.2	0.11
Gastroscopy	0.25	0.08	0.05
Echocardiography	0.3	0.05	0.02

Whether a patient enters the hospital via an emergency room or an outpatient room, s/he is likely to encounter many queues. To begin with, when there is no bed available for a patient, the patient's waiting time increases. Inadequate communication between any two supportive departments in continuation of service to the patient escalates further the patient's waiting time. See Griffiths (2006) about the activities in queue in hospitals. When the patient's waiting time to receive treatment is kept increasing, the workload for the medical professionals would also increase and the quality of services could deteriorate. To counter such impacts, the queuing concepts and tools are exercised. First, the

queuing concepts could help to understand an uncertain demand level for patients' service in a restricted capacity of a hospital/clinic. See Green (2006) for details. To be specific, the queuing methodology helps to decide how many nurses, physicians or support staffs are needed to manage the triage in the emergency or regular units of a hospital based on an estimate of the length of stay by a patient of a type. See Figure 1 and Table 1 to realize how the waiting time for diagnostic test results in an England's hospital occurs, according to Davis (2007), Sanmartin et al. (2000), O'Rourke et al. (2000) and Miller (2004).

This pattern of waiting time is typical in hospitals everywhere. Also, the service time pattern could be understood and improved with an application of queuing concepts. See Lane et al. (2000) for the necessity of improvements in hospital waiting time. An optimal decision could be made with the help of queuing theory on whether the capacity of a unit needs to be expanded. Suppose that in a steady state manner, the patients arrive, form a queue, receive service and then exit a hospital or health clinic system. Lesser the waiting time requires higher cost to the hospital as it requires more resources though the patients save time with more level of satisfaction. This is recognized in the waiting time cost function. On the contrary, the hospital utilizes more resources with an intention to provide lesser waiting time for patients. If the hospital is idle with not enough patients flow, there is going to be more cost to the hospital. This cost is recognized as the idle cost function. But, the sum of the idle and waiting cost functions is the total cost function. The total cost function is convex with a minimal cost value and it is in one-to-one relation with an optimal operational policy for the hospital. The queuing concepts and tools are therefore instrumental to identify and formulate such an optimal policy.

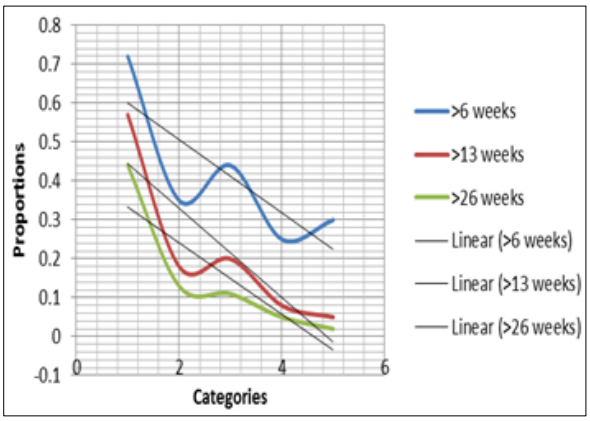

Figure 1: The proportion of patients waiting 6, 13 or 26 weeks: waiting lines.

Besides waiting and idle time costs, more waiting time translates into frustrations and inconveniences to patients. Of course, patient's perception and expectations of hospital services integrate into their frustration level. The frustration level can be reduced only with an application of proper queuing concepts. When a patient is frustrated, it is not good for the hospital management. Why is it so? The stock-holders of the hospital would be unwilling to continue the support the hospital. The agencies might refuse to accredit or license the hospital. The queuing methodology is an excellent concept to comprehend any current status of a hospital operation with an intension to improve it for higher patients' satisfaction which is linked to recruitment of much needed medical professionals, their scheduling, conducive working environment, more productivity, and lesser patient's waiting time with efficient service time. Our discussions below are not limited to physicians, nurses or support staff's best service to patient.

In pharmacy also, the queuing methodologies are useful to address prescription filling time and patient's counseling time. An efficient operation of a hospital system with a lesser queue length and/or minimal waiting time requires optimal strategies based on a comprehensive and correct understanding of the system. The Time Unit (TU) could be hour, day or others but our discussions remain valid. In a non-hospital setting, the nomenclature "patients" is replaced by "customers". An appropriate model needs to be first selected to practice the queuing concepts.

Table 2: Notations in queuing theory.

Notation	Meaning
A (t)	Inter-arrival time distribution
B	Random busy time
B	Erlong's loss (servers are busy)
C	Delay occurs as servers are busy
E (t)	Service time distribution
$E(N_q)$	Expected number of patients in queue
FCFS	First come first served
K	Capacity of the hospital
LCFS	Lost come first served
λ	Arrival rate
μ	Service rate
$\rho = \frac{\lambda}{c\mu}$	Utility level where c is the number of servers
$E(S)$	Expected service time
W	Waiting time for a patient in the hospital from entry to exit
W_q	Waiting time for a patient in queue

What is model? The model is an abstraction of the reality. See Mackay et al. (2005) for visual modeling of queuing system in hospitals. The model for queuing concept application was formalized by Kendall (1951). In Kendall's notation, the arrival pattern is first noted. The arrival patterns A (t) around time t could be deterministic (D) or stochastic (Table 2).

Under stochastic scenario, the arrival pattern is general (G), Poisson (P) or another type. What is Poisson trend? An integer random variable, X is thought to follow a Poisson probability pattern with a rate $\lambda > 0$, where $\lambda = E(X)$ and probability mass function

$$f(x|\lambda) = \lambda^x e^{-\lambda} / [x(x-1)(x-2)....1];$$
$$x = 0,1,2,...,\infty; \lambda > 0$$

Secondly noted is the service time pattern S (t) around time t and it is either constant (C) or stochastic. Under stochastic scenario, the service pattern is general (G), exponential (E) or another pattern.

What is an exponential pattern? A continuous random variable, X is thought to follow an exponential probability pattern with a rate $\lambda > 0$, where $\lambda = 1/E(X)$ and probability density function $f(x|\lambda) = \lambda e^{-x\lambda}; x \geq 0$.

The waiting time for hospital service is significantly reduced by induction of more physicians or increasing the number of hospital beds. Hence, thirdly noted is the number of servers (c) providing service to the patients. Because, the number of servers, $c \geq 1$ providing service makes a difference in the queue length or a patient's waiting time. Most common is a single server (that is, a single physician) for a patient's service. A patient might have to undergo several servers and queues due to registration, to be checked by a nurse, do paper work with casualty officer, take x-ray etc. Another aspect which determines the queue length or waiting time is the capacity (CAP) of the waiting room in a hospital. Fourthly noted is the capacity. Unless specified, the capacity is considered infinite but is finite in reality. When the capacity is full, the entry to the system is denied. The capacity (K) of the waiting room makes a difference for a patient to enter or leave the hospital system. In some situations, the capacity might be infinite type. Fifthly noted is the queue discipline and it makes a difference in the waiting time of a patient. The waiting time of a patient depends on how many patients are prior in the queue to be picked up for their service. The queue discipline is the rule to pick a patient for service. The queue discipline might be first come first served (FCFS), last come first served (LCFS) or a preemptive type. For an illustration, the Kendall's notation $P/E/s/K/FCFS$ embodies the Poisson arrival pattern, exponential service time pattern, s number of servers, a finite capacity K in waiting room and the service is given to patients on first come first basis.

How many patients arrive per Time Unit (TU) is recognized as the arrival rate $\lambda > 0$. What is arrival rate? The number of customers on the average entering a system for service in a defined unit of time is recognized as arrival rate. It might be a Poisson arrival. What is Poisson arrival? In a Poisson arrival pattern, the probability for a single patient to arrive into the hospital in a TU is $\lambda * TU$ which is $1 - p_0$ where p_0 is the probability of no patient entering into the hospital in the TU. In other words, the probability for two or more patients arriving in TU is negligible in a Poisson arrival pattern with rate.

What is service discipline? The manner in which a customer is selected from a pool is recognized as service discipline. When the first arriving customer gets selected for service, it is called First Come First Served (FCFS). When the last arriving customer is selected for service, it is called Last Come First Served (LCFS). When a particular customer is selected over others for some reason to receive a service, it is recognized as priority service.

What is service rate? The number of customers getting the full service from one or more servers in a defined unit of time on the average is recognized as service rate. Suppose that the service rate is $\mu > 0$ per TU and the service follows an exponential probability pattern. What is memory less? A system with uncertain input or output is thought to function with no memory if its conditional probability statement

$P(X > x + m) = P(X > m); x \geq 0; m \geq 0$ is valid. The exponential probability distribution is memory less in the sense that the probability for a patient to take more than *t* units of time is same irrespective whether it occurs in the beginning, end or any part of the hospital operation.

The ratio of arrival rate to the service rate is a crucial factor to the system's *utility level* and it is indicated by $\rho = \frac{\lambda}{\mu} \geq 0$. When $\rho \geq 1$, the system is considered unstable and the queue gets out of bound. It is meaningless to discuss such a scenario. The concept of queue length (L_q) or waiting time (W_q) (in queue is meaningful only when $\rho < 1$. In a general service time pattern, the expected waiting time, E (W) is proportional to the expected service time E (S), utility level as shown in the *Pollaczek-Khintchine law* (1) below.

$$E(W) = \frac{\rho[E(S)]^2 + Var(S)}{2E(S)[1-\rho]} \quad . \qquad (1)$$

According to the law (1), an increase of the utility level to near one will result in a tremendous increase of patient's waiting time. Several adjustments to hospital system

could decrease the patient's waiting time. For an example, like shifting a nurse in a clinical unit to blood pressure measurement unit or the front desk might reduce the queue length. The scheduling patients on appointment basis, regrouping employees to increase the patient flows etc. are some strategies. In developed and developing nations, a topic for extensive discussions and debates is the importance of reducing the total healthcare cost. The average service cost is consistently decreasing, when the hospital capacity increases. The waiting line cost is monotonically increasing step by step gradually as the capacity is expanded. Let the sum of the service cost and the waiting line cost is indicated by the total cost. The total cost is decreasing in the beginning up to a point and then increases gradually. The total cost reaches an optimal low minimal value when the hospital capacity increases and it is the optimal value. The optimal hospital capacity, optimal service cost and optimal waiting time could be identified based on the total cost curve.

A well-known balancing equation is $L_q = \lambda W_q$ and it is recognized as *Little's queue law*. What is Little's law? The relationship between the queue length, L and the waiting time, W of a customer is recognized as *Little's law* $L = \bar{\lambda} W$ where $\bar{\lambda}$ is the average number of customers entering the queuing system. The number of patients waiting in the hospital system is $W = \frac{1}{\mu} + W_q$ where $\frac{1}{\mu}$ denotes the number of customers receiving the service. A summary of these different situations along with queue length, patients receiving service, expected waiting time of a patient and his/her expected service time are tabulated below for a comprehension. The Table 3 summarizes the number of patients waiting in line, the number of patients receiving service, the waiting time in line and the service for patients in a hospital under various queue disciplines.

IS A HOSPITAL BUSY OR IDLE AT A GIVEN TIME?

A hospital's operation period is a mix of idle and busy durations. The busy duration starts with an arrival and ends when no patient is yet to be serviced. A busy cycle is the time between two successive arrivals of an idle duration and it is also the sum of a busy and an adjacent idle periods. For an example, under a single server queuing scenario: $P/E/1/\infty/FCFS$, the hospital's expected busy time is

$$E(BusyTime) = [\mu(1-\rho)]^{-1} \tag{2}$$

with a variance of

$$Var(BusyTime) = [\frac{(1+\rho)}{(1-\rho)}][E(BusyTime)]^2 . \tag{3}$$

What is busy time? The busy time is an average time, ρ during the servers in a system is busy taking care of customers and it is calculated using $\rho = \frac{arrival_rate}{service_rate}$. Realizing that the expected waiting time for a patient increases when the service rate decreases. A system is more volatile when the variance is larger. The hospital system becomes more volatile when the expected waiting time for a patient is higher.

Table 3: Summary of hospital scenarios.

Situation	Patients in queue L_q	Patients receiving service L-L_q	Waiting time in queue W_q	Service time W-W_q
$P/E/1/\infty/FCFS$	$\frac{\rho^2}{1-\rho}$	ρ	$\frac{\rho}{\mu(1-\rho)}$	$\frac{1}{\mu}$
$P/E/s/\infty/FCFS$	$\frac{p_0(s\rho)^s\rho}{s!(1-\rho)^2}$	$s\rho$	$\frac{L_q}{s\mu\rho}$	$s\rho$
$P/E/s/K/FCFS$	$\frac{p_0(s\rho)^s\rho}{s!(1-\rho)^2}[1-\rho^{K-s}-(K-s)\rho^{K-s}(1-\rho)]$	$s\rho[1-\frac{p_0\rho}{s!}\frac{(s\rho)^K}{s^{K-s}}]$	$\frac{L_q}{\mu(L-L_q)}$	$\frac{1}{\mu}$
$P/G/1/\infty/FCFS$	$\frac{\lambda^2(\sigma^2_{service}+\mu^2_{service})}{2(1-\rho)}$	ρ	$\frac{L_q}{\lambda}$	$\frac{1}{\mu}$

When the expected waiting time is longer, the patients might be lost forever or re-enter the hospital after several retrying. See Artalejo (2010), Nosek et al. (2001), and Shin et al. (2009) for details about the ideas of leaving and reentering the queue. See Harper et al. (2003). for ways to decrease a queue's length. Assuming that the time between retrials follows a memory less exponential distribution with parameter η, the number of patients in the new queue is

$$L_q = \frac{\rho^2}{1-\rho} + \frac{\lambda\rho}{\eta(1-\rho)} \tag{4}$$

with a waiting time

$$W_q = \frac{\lambda(\eta\rho+\lambda)}{\eta(1-\rho)} . \tag{5}$$

NEW IDEAS FOR HOSPITALS TO BE ALERT WHEN IMPATIENT PATIENTS EXIS

In this section, new ideas are introduced to advance the state of the art in queuing theory to suit the scenario in which impatient customers exist along with alert servers in the system. See Holt et al. (2011) for tackling the impatient customers and their impact on the waiting time. Patience is one of seven noble virtues in life. But, in a healthcare hospital setting, the patience should be assessed as the state of mental endurance in an uncomfortable condition. Impatience is a consequence of service delay in comparison to an expectation level. In a normal queuing scenario, the first come patient is first served. Even in such normal scenario, a patient might not join the queue after noticing a longer queue. Otherwise, depending on the severity of illness, the patients in a waiting line might exhibit quite impatience. In the midst of an impatience level d, the arrival rate could alter to $\lambda^* = \lambda / d$. In a normal scenario, the impatience level is $d = 1$ and it is the baseline level in the discussion so far. The impatience level is determined in a scenario by $d = \frac{L_q}{m_{\min} - L_q}$ where m_{min} is a minimum acceptable number of patients in the queue for a new arriving patient. Note that $d = 1$ corresponds to $L_q = m_{\min} / 2$. When $L_q > m_{\min} / 2$, note that $d > 1$ and consequently, the arrival rate is deflated. When $L_q < m_{\min} / 2$, note that $d < 1$ and consequently, the arrival rate is inflated.

Independently of how patients make choices, the hospital units might exercise alert levels on their service rate. In a normal operation, the service is at a baseline level which is indicated by an alert level r = 1. When the hospital realizes that number, L_q of patients waiting in queue is excessive of a maximum anticipated level $m_{\max}$, the service rate is increased to $\lambda^* = r\mu$ with an increased alert level, $r \geq 1$, where $r = \frac{L_q}{m_{\max} - L_q}$. In a contrary scenario, the hospital might opt to reduce the number of patients to be serviced and it corresponds with $0 < r < 1$. The service rate is altered to $\mu^* = r\mu$ depending on the chosen alert level in the hospital who serves the patients. Note that $r = 1$ corresponds to $L_q = m_{\max} / 2$. When $L_q > m_{\max} / 2$, note that $r > 1$ and consequently, the service rate is increased. When $L_q < m_{\max} / 2$, note that $r < 1$ and consequently, the service rate is decreased.

In this scenario where the patients are impatient and the hospital operates under a chosen alert level, the hospital's new utility level becomes $\rho^* = (\frac{1}{rd})\rho$ with a restriction $\rho < rd$. The balance between the normal and altered scenarios is controlled by $\frac{1}{rd}$. When $rd < 1$, the hospital is more utilized than in a normal scenario. Otherwise, the hospital is under-utilized.

Consequently, under a single server queuing scenario: $P/E/1/\infty/FCFS$, the hospital's expected busy time in an alert hospital operation amidst impatient patients is

$$E(BusyTime) = [\mu(r - \frac{\rho}{d})]^{-1} \tag{6}$$

with a variance of

$$Var(BusyTime) = [\frac{(rd+\rho)}{(rd-\rho)}][E(BusyTime)]^2 \tag{7}$$

Also, after several leaving and reentering the hospital, the number of patients in the new queue is

$$L_q = \frac{\rho^2}{rd(rd-\rho)} + \frac{\lambda\rho}{\eta d(rd-\rho)} \tag{8}$$

with the waiting time

$$W_q = \frac{\lambda(\eta\rho + r\lambda)}{\eta d(rd-\rho)} . \tag{9}$$

ANALYSES OF PATIENTS" WAITING AND SERVICE TIMES

In this section, the above mentioned queuing concepts and tools are narrated in the context of a hospital data in Malta. Malta is a largest island of an archipelago situated in the center of the Mediterranean with a total population of a million. There is one acute general teaching hospital [Mater Dei Hospital - MDH) offering a range of hospital services at no cost, under its National Health Service system. This is the only center in Malta providing critical care services. Several Health Centers are available all around the Island providing non-critical, elective as well as acute care, at no cost. Prolonged waiting times at the MDH Emergency Department (ED) have been heavily criticized by the public. The prolonged waiting times triggered public dissatisfaction with their services, and patients are known to leave without waiting to be seen. Timely patient care was mandated to improve patient health outcomes. Upon arrival and registration at MDH, the patients are assessed in three triage categories: life threatening (priority one), urgent (priority two), and non-urgent (priority three). The latter are patients that could have been seen by a family doctor outside the hospital premises.

The data of this section were collected as follows. A study (see Azzopardi et al. 2011, for data) was conducted with an aim to reduce the waiting times in all three priority wards during four shifts of six hours each. The first shift is from 8 AM to 2 PM. The second shift is from 2 PM to 8 PM. The third shift is from 8 PM to 2 AM. The fourth shift is from 2 AM to 8 AM. The three triage areas were constantly monitored, 24 hours a day, for a period of one week. Factors contributing to patient care delay were identified to significantly decrease length of stay and thus improve the patient care efficiency in the emergency department. Data on eighteen variables were collected and the variables are: Y1 (waiting time to enter triage), Y2 (triage time taken), Y3 (waiting time to enter area), Y4 (waiting time from registration to first assessment), Y5 (waiting time to be seen by nurse), Y6 (total interactive time with nurse), Y7 (waiting time to be seen by a casualty officer), Y8 (total interaction time with a casualty officer), Y9 (time from first seen to last seen), Y10 (waiting time by senior review), Y11 (total time to take x-ray), Y12 (total time for computed tomography), Y13 (total time for taking ultra sound test), Y14 (total treatment time in the ward), Y15 (waiting time to leave the emergency ward after admission), Y16 (waiting time to leave the emergency ward after discharge), Y17 (total time in the hospital area) and Y18 (total time in emergency department). To be brief, only the variables Y5, Y6, Y7 and Y8 are selected here to illustrate the queuing concepts, the existence of different patient's impatience levels and the alert levels of the wards during the four shifts. See Figures (7 through 9) to compare the three wards and the shifts within a ward. The waiting time, Y5 to be seen by a nurse and hence the impatience level are more in 3rd shift across all three wards. The three wards are at different alert levels. The interactive time with a nurse, Y6 is more in shift 2 in all three priority wards but is longer only in priority ward 1 (Figure 2).

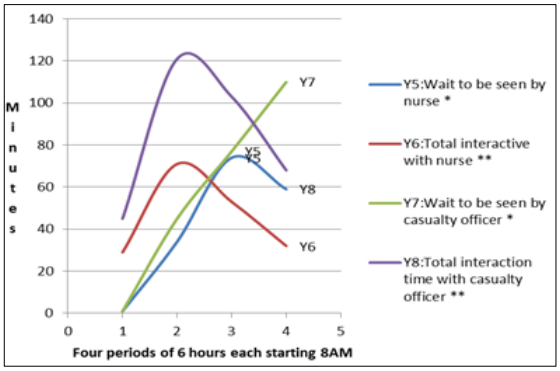

Figure 2: Comparison of periods in emergency ward 1 for life threatening cases.

The waiting time to be seen by a casualty officer, Y7 increases monotonically over the shifts in priority ward 1 but gets better in priority wards 2 and 3 proving that the patients' impatience level changes differently among the three priority wards.

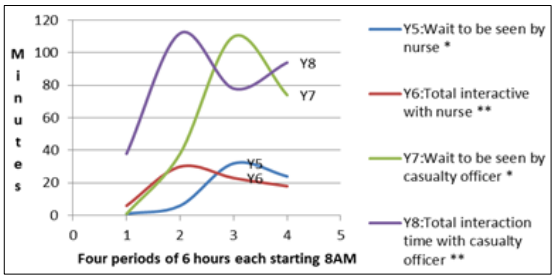

Figure 3: Comparison of periods in emergency ward 2 for urgent cases.

The total interaction time with the casualty officer, Y8 is more in shift 2 in all three priority wards but lesser in shift 3 (Figure 3) only in priority ward 2 (Figure 4). In all three priority wards, the casualty officer takes more time with the patients. These are providing evidence for the

existence of patients' impatience level and the hospital's alert level.

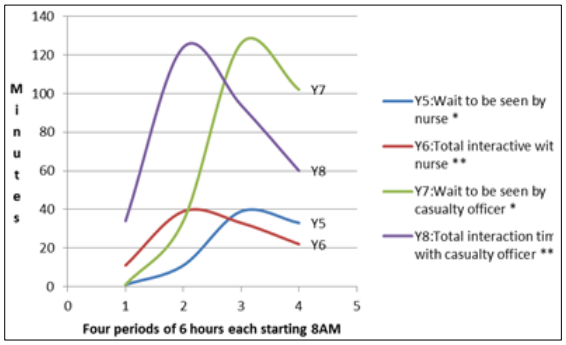

Figure 4: Comparison of periods in emergency ward 3 for non-urgent cases.

FUTURE RESEARCH DIRECTIONS

The theoretical results of our new model pave a way to analyze data of the type in the previous section. The finding indicates that the patterns are quite different among the life threatening cases, urgent but not life threatening cases and the non-urgent cases. The findings out of the analyzed data for the life threatening cases reveal important practical knowledge. Most of the patient's waiting time is spent for to be seen by the nurses. Whether the time could be reduced by increasing the number of nurses? An answer depends on the steady state of the incoming patients flow? A study must be done to decide the optimal number of available nurses depending on the patient's arrival pattern. The next big chunk of waiting time is for interaction with the nurses and it is crucial for the patient. The time spent by a patient to be seen by the casualty office increases over the four periods of shift time meaning that the time is lesser in the first shift but more in later shifts. Our new model detects this realistic meaningful practicality. The life threatening cases are too important and serious to be ignored or rushed through by the nurses and/or the casualty officer. This finding is very practical.

With respect to the urgent but not life threatening cases, the findings are different but practical again. The total interaction time with the nurse and casualty officer in the hospital is longer in the second shift, lower in the third shift but more again in the fourth shift. Interestingly, the time spent with the casualty office is more in the third shift but lesser in the fourth shift and this finding is quite opposite of the total time spent by the patient. Our new model captures this opposite finding. The time spent by a nonlife threatening urgent patient with the nurse is quite parallel but not the same as by a life threatening patient.

The total time spent by a non-urgent patient with nurse and casualty officer is more in the second shift and lesser and lesser in the third and fourth shifts in parallel pattern to that of life threatening patient. Recall that the total time spent by a non-life threatening urgent patient has oscillated with a minimal occurring in the third shift. The time spent with a casualty office by a non-urgent or non-life threatening urgent patient is quite parallel to each other but is quite different from that of life-threatening patient. The interactive time with a nurse by patient is parallel in all three types: life-threatening cases, non-life threatening but urgent cases and non-urgent cases. Very strikingly, the patients in all three groups wait longer to be seen by the nurse in shift three and lesser time in shifts two or four.

COMMENTS AND CONCLUSIONS

Given the necessity of long waited healthcare reforms in developing and developed nations, the future research directions need to configure the human elements of patients in hospitals/clinics. In 21st century of globalized thinking and living choices, the medical tourism is going to flourish. The eventuality of standardized healthcare across developing nations as much in developed nations needs to be thoroughly researched out. The impediments and the remedies of such healthcare standardization have to be worked out. The demand for patients care in developing nations is perpetually higher than the capacity to serve the patients. The patients are obligated to discharge their duties [Evans, 2007]. The patient's duties are of paramount importance especially in limited healthcare resources. The patients have to seek and access the resources responsively. Otherwise, the queue is likely to grow unboundedly in hospitals/clinics.

Reasons for growing queues include but are not restricted to overcrowded waiting rooms, lack of beds, patients occupying beds without medical receiving treatment due to lack of trained physicians, etc. Managing length of stay is a challenging issue for almost all hospitals. The allocation of surgical theaters among the specialists (such as cardiac, orthopedic, neurologic, urologic etc.) warrants research attention. The causes for these reasons are currently not well understood. An equilibrium level needs to be established to accommodate the patients who leave without being seen by the physicians. The queue concepts and methodologies are integral and important parts of such future research directions. The findings, new model and methodology are limited to the validity of the assumption of the steady state incoming and outgoing patients' flow patterns.

Funding: No funding sources
Conflict of interest: None declared
Ethical approval: The study was approved by the institutional ethics committee

REFERENCES

1. Artalejo, JR. Accessible bibliography on retrial queues: progress 2000-2009. Math Comp Mod. 2010;51:1071-81.
2. Azzopardi M, Cauchi M, Cutajar, K, Ellul R, Mallia-Azzopardi C, Grech V. A time and motion

study of patients presenting at the accident and emergency department at Mater Dei hospital. BMC Res Notes. 2011;4:1-11.
3. Buhaug H. Long waiting lists in hospitals: Operations research needs to be used more often and may provide answers. Br Med J. 2002;324, 252-253.
4. Davis A. Thousands of patients wait more than 26 weeks for tests. Br Med J. 2007;335:365.
5. de Bruin AM, van Rossum AC, Visser MC, Koole GM. Modeling the emergency cardiac in-patient flow: an application of queuing theory. Health Care Manag Sci. 2007;10:125-37.
6. Dunnill MG, Pounder RE. Medical outpatients: changes that can benefit patients. Clin Med. 2004;4:45-9.
7. Evans H. Do patients have duties? J Med Ethics. 2007;33:1.
8. Green L. Queuing analysis in healthcare. In: Hall RW, eds. Patient Flow: Reducing Delay in Healthcare Delivery. 1st ed. New York, Springer Press; 2006: 281-308.
9. Green LV, Soares J, Giglio JF, Green RA. Using queuing theory to increase the effectiveness of emergency department provider staffing. Acad Emerg Med J. 2005;13(1):61-8.
10. Griffiths JD, Lloyd NP, Smithies M, Williams J. A queuing models of activities in an intensive care unit. IMA J Manag Math. 2006;57(6):711-7.
11. Harper PR, Gamlin HM. Reduced outpatient waiting times with improved appointment scheduling: a simulation modeling approach. OR Spect. 2003;25:207-22.
12. Holt N, Smucker D, Hester J. Impact of an inpatient palliative care consult service on length of stay and urgent patient resuscitations in the ICU (727). J Pain Symp Manag. 2011;41(1):288.
13. Jacobs R. Alternative methods to examine hospital efficiency: data envelopment analysis and stochastic frontier analysis. Health Care Manag Sci. 2001;4:103-15.
14. Kampllikar M. Losing wait. The TMTC J Manag. 2005:31-5.
15. Kennedy J, Rhodes K, Walls CA, Asplin BR. Access to emergency care: restricted by long waiting times and cost and coverage concerns. Ann Emerg Med. 2004;43:567-73.
16. Lane DC, Monefeldt C, Rosenhead JV. Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department. J Operat Res Soc. 2000;51(5):518-31.
17. Mackay M, Lee M. Choice of models for analysis and forecasting of hospital beds. Health Care Manag Sci. 2005;8:221-30.
18. Matthias A, Marija C, Cutajar K, Ellul R, Mallia-Azzopardi C, Grech V. A time and motion study of patients presenting at the accident and emergency department at Mater Dei Hospital. BMC Res Notes. 2011;4:421.
19. Miller GG. Waiting for an operation: patients' perspectives. Canadian J Surg. 2004;47(3):179-81.
20. Nosek Jr R, Wilson J. Queuing theory and customer satisfaction: a review of terminology, trends, and applications to pharmacy practice. Hosp Pharm. 2001;36:275-9.
21. O'Rourke N, Edwards R. Lung cancer waiting times and tumour growth. Clin Oncol. 2000;12:141-4.
22. Ostendorf M, Buskens E, van Stel H. Waiting for total hip arthroplasty. Avoidable loss in quality time and preventable deterioration. J Arthroplas. 2004;19:302-9.
23. Preater J. Queues in health. Health Care Manage Sci. 2002;5:283-93.
24. Propper C, Croxson B, Shearer A. Waiting for hospital admissions: the impact of fund bolding. J Health Econom. 2002;21(2):227-52.
25. Sanmartin C, Short SE, Barer ML, Sheps S, Lewis S, McDonald PW. Waiting for medical services in Canada: lots of heat, but little light. Canadian Med Assoc J. 2000;162(9):1305-10.
26. Sternberg S. Hospitals too slow on heart attacks: study finds only 35% react quickly enough, 2006. Available at: http://usatoday30.usatoday.com/news/health/2006-11-12-angioplasty_x.htm.
27. Shin YW, Choo TS. M/M/s queue with impatient customers and retrials. Appl Math Mod. 2009;33:2596-606.
28. Tucker JB, Barone JE, Cecere, RG, Blabey CK, Rha R. Using queuing theory to determine operating room staffing needs. J Trauma Injury Infect Crit Care. 1998;45(1):192.
29. van Klei WA, Moons KG, Rutten CL, Schuurhuis A, Knape JT, Kalkman CJ, et al. The effect of outpatient preoperative evaluation of hospital inpatients on cancellation of surgery and length of hospital stay. Anesthesiol Analog. 2002;94:644-9.
30. Vericourt F, Jennings O. Nurse staffing in medical units: a queuing perspective. Operat Res. 2011;59(6):1320-31.
31. Wellstood K, Wilson K, Eyles J. Unless you went in with your head under your arm: Patient perceptions of emergency room visits. Soc Sci Med. 2005;61(11):2363-73.

DOI: 10.5455/2320-6012.ijrms20140882
Cite this article as: Shanmugam R. How do queuing concepts and tools help to efficiently manage hospitals when the patients are impatient? A demonstration. Int J Res Med Sci 2014;2:1076-84.