Accepted Manuscript

Application of queueing theory in health care: A literature review

C. Lakshmi, Appa Iyer Sivakumar

PII: S2211-6923(13)00003-9
DOI: http://dx.doi.org/10.1016/j.orhc.2013.03.002
Reference: ORHC 18

To appear in: *Operations Research for Health Care*

Received date: 3 July 2012
Accepted date: 4 March 2013

Please cite this article as: C. Lakshmi, A.I. Sivakumar, Application of queueing theory in health care: A literature review, *Operations Research for Health Care* (2013), http://dx.doi.org/10.1016/j.orhc.2013.03.002

Application of queueing theory in health care: A literature review

C. Lakshmi[1], Appa Iyer Sivakumar[2]

Abstract

This paper reviews the contributions and applications of queueing theory in the field of health care management problems. This review proposes a system of classification of health care topic areas examined with the assistance of queueing models. The categories described in the literature are expanded and a detailed taxonomy for sub-groups are formulated. The goal is to provide sufficient information to analysts who are interested in using queueing theory to model a health care process and want to locate the details of relevant models.

Keywords Queueing theory, literature review, health care

1 Introduction

Health care managers face the challenging task to organise their processes more effectively and efficiently. The pressure on health care managers rises as both demand for health care and expenditures are increasing steadily. Health care is an enormous field and there are many different ways of analysing it. How to decide on the best locations of medical clinics and emergency vehicles for providing maximum health care coverage to a given population? How many base locations of medical ambulances are needed if the total distance from the locations to the hospitals must be less than a given number? How should radiation treatment be planned for minimising the treatment time of a cancer patient? How should the nurses in a trauma center be scheduled and re-rostered to maintain an adequate service level even in the worst-case scenario? Many problems like these need to be addressed in health care, and Operations Research provides numerous methodologies and solution techniques for tackling them. The operational research model offers a systematic approach to problem solving and allows for the characterisation of activities of an existing system using mathematical modelling. One approach, based on mathematical models, that successfully addresses problems in the health care system is queueing models .

A review of the literature was conducted to determine the up-to-date health care areas supported by queueing models. We looked for articles that described qeueing models and their topics, or their keywords, related to health care or population health issues. This paper focuses on selected research articles published in the field of queueing theory applications (QTA) in Health Care Management (HCM) and the articles published between 1952 and 2011. The aim of this paper is to describe the main trends in the application of queueing models that are available to health care decision makers, while bearing in mind that the resulting bibliography should not be regarded as the final one. In this rapidly developing field, new articles are constantly appearing in scientific journals and new examples of applications waiting to be included in the list.

The remainder of this paper is organized as follows: Section 2 describes the search strategy

[1]Singapore-MIT Alliance, N3.2-01-36, 65, Nanyang Drive, Singapore 637460, lakmaths@gmail.com
2 School of Mechanical and Aerospace Engineering, Nanyang Technological University, Singapore 637460, msiva@ntu.edu.sg

and selection of papers. In section 3, we provide the taxonomy of literature on applications of queueing theory in health care. Classification methods on QTA research in HCM are given in section 4. Section 5 includes the list of literature on simulation-based queueing model in health care. In section 6, a describtive analysis on the review is presented. Finally, section 7 presents the summary and conclusions.

2 Search strategy and selection of papers

In our search we included the ACM Digital Library, EBSCO databases (Medline, Business Source Complete, Academic Search Complete), Elsevier and Springer journals (Science Direct, Medline, Inspec), INFORMS publications journals and conference proceedings. The search strategy consisted of three stages. In the initial stage, digital libraries were scanned. A query was formulated to identify the papers with words queueing models in the title or in the keywords and one of the words health, hospital, patient, emergency department, outpatient, surgery, ambulance, pharmacy, epidemic and disaster in the abstract. During the second stage, the decision about paper inclusion was made, based on the following criteria: (1) the paper describes a queueing model and its applications in health care management; and (2) a clear association with a general phenomenon of health care management issues. The final stage involved manual scanning of papers from the conference proceedings from 2000-2011. The inclusion criteria remains the same as for journal publications.

After the collection of the initial set of papers, we began to assign them according to decisions reached regarding the area of application and the modelling method. The reviewed articles are (directly or indirectly) related to queueing theory applications research in healthcare . The literature review applications found 141 article on queueing theory research in health care reported (during 1952 to 2011) in various outlets such as journals and other publications outlets, as shown in Table 1.

List	1952-1959	1960-1969	1970-1979	1980-1989	1990-1999	2000-2011	Total
Journals	2	3	11	12	21	63	112
Proceedings					1	16	17
Lecture Notes						1	1
Technical Reports		2	1			5	8
Working papers						2	2
Books						1	1
Total	2	5	12	12	22	88	141

Table 1: Number of manuscripts in the original set, categorised according to publication type and publication year

As seen from Table 1, this set largely consists of articles published in scientific journals. Note that almost half of the contributions appeared in or after 2000, which illustrates the increasing interest of researchers in this domain. Since the total number of manuscripts is large and our main interest is directed towards the recent advances proposed by the scientific community, we restrict the set of manuscripts to those published in or after 2000. This is because of advancement in computational power and software availability. We also limited the contributions that are incorporated in this review to those that are written in English in order to augment the paper's accessibility.

3 Application area of queueing theory in health care

Operation research existed as a scientific discipline since 1930's. It is a discipline of applying appropriate analytical methods for decision making. Many optimisation techniques such as linear and dynamic programming models allow quantitative analysis; these mathematical models are generally neglecting the effects of uncertainty and assume that the results of decisions are predictable and deterministic. This abstraction of reality allows large and complex decision problems to be modelled and solved using powerful computational methods. Simulation models allow for detailed representation of complex, dynamic behaviour; however, building large-scale hospital simulation models is not easy. One of the major issues in using simulation is the extensive data requirements that are needed to support such studies and the prohibitive expenses associated with such data collection. Also, selection of the level of detail (in the model) depends strongly on the focus of the problem to be addressed One of the major uses of operational research in healthcare is in the form of queueing theory. We believe that queueing has the advantage of producing simple models using less data while including randomness [see, Cochran et al (2006a)].

Queues or queueing theory was first analysed by A.K Erlang in 1913 in the context of telephone facilities. It is extensively practiced or utilised in industrial setting or retail sector-operations management, and falls under the purview of decision Sciences. The rising cost of health care can be attributed not only to ageing population and new expensive and advanced treatment modalities but also to inefficiencies in health delivery. Queueing theory application is an attempt to minimise the cost through minimisation of inefficiencies and delays in the system .There are many problems in health care system which can be solved using queueing theory in operational research.

Queueing theory (for more details, see Gross and Harris (1998)) constitutes a very powerful tool because queueing models require relatively little data and are simple and fast to use. Because of this simplicity and speed, modellers can be used to quickly evaluate and compare various alternatives for providing service. Beyond the most basic issue of determining how much capacity is needed to achieve a specified service standard, queueing models can also be useful in gaining insights on the appropriate degree of specialisation or flexibility to use in organising resources, or on the impact of various priority schemes for determining service order among patients. With the ever increasing power of computers there is increasing scope for numerical methods and simulation models to be used alongside traditional queueing theory to help `understand real life queueing systems as well as possible'.

In practice, queue modelling often goes either the analytic route or the simulation route because of the background of the modeller. Each can be inadequate on its own, with the analytic approach requiring unconvincing assumptions to be made, or the simulation approach risking over fitting or development of a sledgehammer to crack a nut. The possibility that numerical methods and simulation can be used 'like formulae' is also clearly very attractive in a practical situation. The importance of integrating analytic and simulation approaches in research and in applications makes a strong case for a new discipline of 'queue modelling'. It could draw upon the combined strengths of analytic and simulation approaches to tackle queueing problems in research and in practice. Alongside these modelling approaches it could take up the challenge of understanding these systems from both the modelling and management perspectives, including the responsibility to draw out managerial insights, all with the purpose ‘to understand real life queueing systems as well as possible.

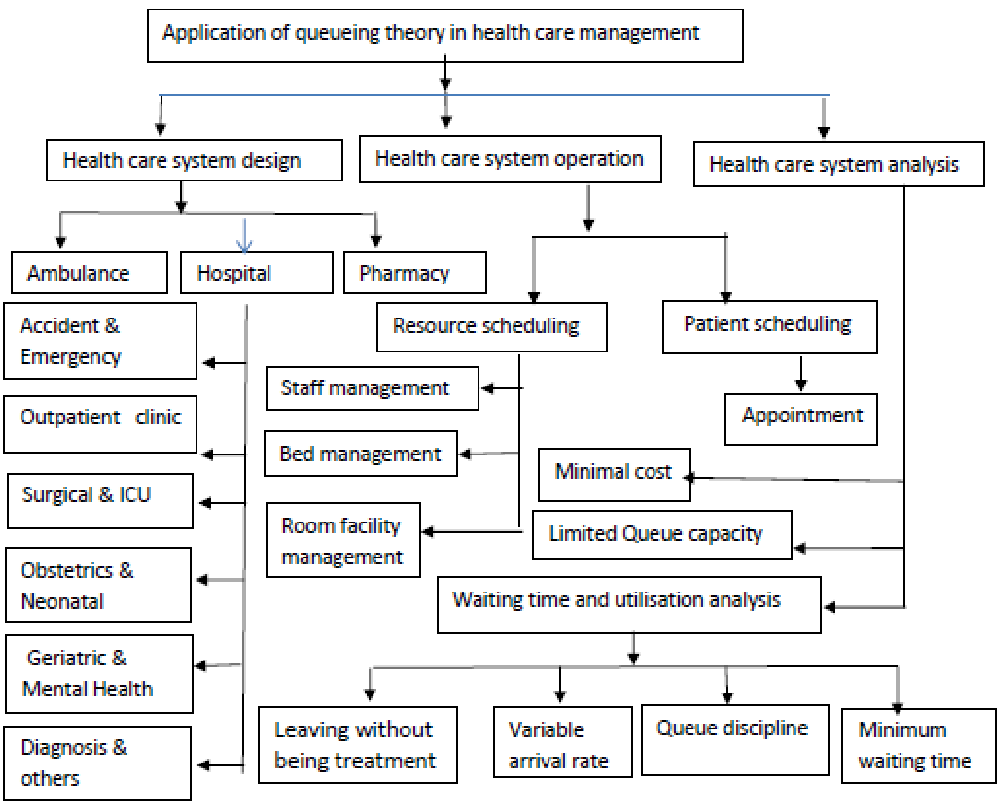
A considerable body of research has shown that queueing theory can be useful in health care and some reviews of this work have appeared. McClain (1976) reviews research on models for evaluating the impact of bed assignment policies on utilisation, waiting time, and the probability of turning away patients. Nearly a decade ago, Preater (2001) compiled a bibliography of queueing applications in health care. Also Nosek et al. (2001) review the use of queueing theory in pharmacy applications with particular attention to improving customer satisfaction. Customer satisfaction is improved by predicting and reducing waiting times and adjusting staffing. Preater (2002) presents a brief history of the use of queueing theory in health care and points to an extensive bibliography of the research that lists many papers (however, it provides no description of the applications or results). Samuel Foumndam et al. (2007) survey the contributions and applications of queueing theory in the field of health care. They summarises a range of queueing theory results in the following areas: waiting time and utilisation analysis, system design, and appointment systems. Also they considered the results for systems at different scales, including individual departments (or units), health care facilities, and regional health care systems. Pajouh et al. (2010) review and categorise the applications of queueing theory in modelling hospital processes. Also the book on health care (see Hall (2006a)) is dedicated to improving health care through reducing the delays experienced by patients. One aspect of this goal is to improve the flow of patients, so that they do not experience unnecessary waits as they flow through a health care system. Another aspect is ensuring that services are closely synchronised with patterns of patient demand. Still another aspect is ensuring that ancillary services, such as house keeping and transportation, are fully coordinate with direct patient care. Past experience shows that effective management of health care delays can produce dramatic improvements in medical outcomes, patient satisfaction, and access to service, while also reducing the cost of health care.

4 Classification methods

The application of queueing models in the analysis of health care systems increasingly accepted by health care decision makers. This is due in part to the large number of successful queueing theory applications in health care studies reported in the literature. Linda Green (2006a) presents the theory of queueing as applied in health care. She discusses the relationship amongst delays, utilisation and the number of servers; the basic $M/M/s$ model, its assumptions and extensions; and the applications of the theory to determine the required number of servers. Vikas Singh (2006) analyses the theory and instances of use of queueing theory in health care organisations around the world and the benefits accrued from the same. A hypothetical simplistic queueing model is also demonstrated in the literature analysis section to illustrate the point. Creemers Stefan et al. (2007a) explain how a queueing model offers an excellent tool to analyse and to improve the performance of health care systems. They discuss differences with the modelling of manufacturing systems and to focus on modelling issues in patient flow. Foster et al. (2010) present the general background of queueing theory and its associated terminology, and how queueing theory can be used to accurately model the hospital system with an example. Reetu Mehandiratta et al. (2011) analyses the theory (queueing) and instances of use of queueing theory in health care organisations around the world and benefits acquired from the same.

This section reviews the large body of queueing models and analysis efforts that have been reported to address health care management problems and provides an up-to-date, comprehensive collection of articles describing these applications. This section focuses on articles that analyses single or multi-facility health care clinics, including outpatient clinics, emergency departments, surgical centers, orthopedics departments, ambulances and pharmacies. Also, it summarises a range of queueing theory results in the following areas: waiting time and utilisation analysis, system design, appointment and system analysis. Patient scheduling, resource scheduling and ambulance services are likely the most extensively referenced management problems in health care. With the modification of several aspects categorisation, we employed a classification of three major subgroups of queueing research models applied in healthcare mangement: (1) Health care system design (2) Health care system operation and (3) Health care system analysis.

Fig 1: Applications of queueing theory in health care management



4.1 Health care systems design

This group of models attempts to make future forecasts by estimating anticipated population demands and subsequently allocating resources as needed. A variety of solutions geared toward distinct medical institutions (hospital, ambulatory care center), or groups of them, are investigated. The planning process is conducted in terms of quantity (optimization of the number of hospital beds) and quality (ambulance deployment to assure a balanced access to emergency system services). Queueing models have been extensively used to model and analyse different health care systems such as hospitals (Linda Green (2010)), pharmaceutical industry (Viswanadham et al.(2001)) and ambulance services (Singer et al. (2008)).

4.1.1 Ambulatory care

In case of a medical emergency, it is very important for ambulance services to reach the site as fast as possible. Patient waiting time in this situation is a key indicator for ambulance system performance. Bell (1969) uses a multiple server queueing model to determine the number of ambulances needed to achieve specified response rates. The hypercube model which is a spatially distributed queueing model based on Markovian approximations has been one of the most popular techniques to model emergency vehicle systems (Larson (1974), Burwell et al. (1972)). Burwell et al. (1973) develop an extension of the hypercube model that contains preference ties and apply the proposed model to the emergency medical system of Greenville County, South Carolina. They conclude that the proposed model could provide good estimates of the emergency system performance when input parameters are accurately specified. Scott et al. (1978) develop a model that predicts the entire distribution of response time, explicitly accounting for the rate and spatial distribution of demand, variable ambulance velocities, and queueing effects. They test the model using data sampled from 3,936 ambulance runs in Houston and achieved close agreement between empirical and predicted distributions of response time. Mendonca et al. (2001) use the hypercube model to evaluate the mean response time of the system to an emergency call in an emergency medical system on a Brazilian highway connecting the cities of Sao Paulo and Rio de Janeiro. Iannoni et al. (2007) use the hypercube model to analyse emergency medical systems, which can receive calls on-line on highways. They consider partial backup, multiple dispatch, different classes of servers and customers and particular dispatching policies in the model. Mateo Restrepo et al. (2009) introduce the two models for the static ambulance deployment problem. The models capture some of the essential queueing dynamics of an emergency medical service system while allowing efficient solution procedures.

4.1.2 Hospital

Vanberkel et al. (2010) review the quantitative health care models to illustrate the extent to which they encompass multiple hospital departments. Timely access to care is a key component of high quality health care. Yet, patient delays are prevalent throughout the health-care system resulting in dissatisfaction and adverse clinical consequences for patients as well as potentially higher costs and wasted capacity for providers. Arguably, the most critical delays for health care are the ones associated with health-care emergencies. Unfortunately, emergency department (ED) overcrowding is a continuing and growing problem. Adrian H. Zai et al. (2009) use a queueing theory model to simulate three different potential solutions to decrease the delay from patient identification to connection with discharge services. Karnon et al. (2009) describe several cases studies that illustrate how modelling and simulation can assist governments to carry out these responsibilities. The examples are set in a variety of contexts, namely, a hospital emergency department, patient flow in an acute hospital, a national screening program for cervical cancer, and aged care. They employ a range of models including, time series, compartmental models, Markov chains and queueing models. Biju et al. (2011) explain how queueing theory can be studied in integration with other aspects of hospital management such as financial human relations, marketing and other aspects.

4.1.3 Pharmacy services

Queueing theory is used extensively in pharmacy operations management. In a pharmacy, queueing theory can be used to assess a multitude of factors such as prescription fill time, patient waiting time, patient counseling-time, and staffing levels. The application of queueing theory may be of particular benefit in pharmacies with high-volume outpatient workloads and/or those that provide multiple points of service. By better understanding queueing theory, service managers can make decisions that increase the satisfaction of all relevant groups-customers, employees, and management.

Donehew et al. (1978) use queueing theory to address prescription queues and work measurement assessment of prescription fill times. Shimshak et al. (1981) consider a pharmacy queueing system with pre-emptive service priority discipline where the arrival of a prescription order suspends the processing of lower priority prescriptions. Different costs are assigned to wait-times for prescriptions of different priorities. Vemuri (1984) uses computer simulation with a queueing model to assess patient waiting time in the outpatient pharmacy at the Medical College of Virginia. This study concluded that the most significant factor contributing to patient waiting times was the interaction between pharmacy service providers, specifically the typist and the technician. Many different mathematical equations can be used to describe queue formation and behavior; however, the decision to choose one over the other is beyond the scope of this article. In a study by Moss (1987), queueing theory is used to assess the relationships among the number of pharmacy staff members, prescription dispensing process, and outpatient waiting times. He uses a mathematical queueing model to estimate the probability of waiting time exceeding a given value, when prescription arrival and service rates and number of servers are known. The study reveals that the major factors determining outpatient waiting time were the arrival pattern of prescriptions at the pharmacy, sequencing of work, and percentage of staff at work. Lin et al. (1996) use workflow analysis and times study to identify factors leading to excessive waiting times in an ambulatory pharmacy at the University Hospital Inc. (TUH), Cincinnati, Ohio. Boyce et al. (1998) seek to determine the impact of a computerised waiting time program on order turnaround time in a hospital pharmacy. Perhaps the most common and useful application of queueing theory in pharmacy operations is to reduce patient waiting time and maximise staff effectiveness. Lin et al. (1999) focuses on work measurement and computer simulation, which are used to assess the re-engineering of community pharmacies to facilitate patient counseling. Although queueing theory is never mentioned in these articles, the authors used many concepts similar to queueing theory and their results could be instrumental in designing queueing applications for reducing patient waiting time and improving staff utilisation. Eugene Day et al. (2010) improve the customer service and optimise productivity of the pharmacy help desk by using a combination of discrete event simulation and queueing statistical analysis. Ndukwe et al. (2011) describe the queue discipline of the outpatient pharmacy, which involves instituting a cross-sectional intervention by streamlining queue behavior and to measure the impact of streamlining queue characteristics and queue discipline on waiting time of patients.

4.2 Health care systems operation

Hospitals and clinics are facing increasing competition for their services. To attract new patients and retain their patronage, hospitals and clinics must be able to provide fast and efficient health care. Effective and efficient patient flow is indicated by high patient throughput, low patient waiting times, a short length of stay at the clinic, and low clinic overtime, while simulataneously maintaining adequate staff utilisation rates and low physician idle times. Two areas that impact patients in clinics are: resource scheduling and patient scheduling.

4.2.1 Resource scheduling

With the rise in the cost of providing quality health care, hospital and clinic administrators are practicing cost containment by minimising resources for health care provisions while still striving to provide quality health care for patients. This predicament is becoming more prevalent in the health care community as indicated by the large body of literature that analyses the allocation of scarce health care resources. The allocation of resources can be divided into three general areas: bed management, staff management and room facility management.

Table 2: Classification scheme: Health care system design and system operation

Dept.	Bed management	Staff management	Room facility management
ED	Cooper et al. (1974) Green (2002) Shmueli et al. (2003) Bruin et al. (2005) Cochran et al. (2006) Bruin et al. (2007)	Agnihothri et al. (1991) Tucker et al. (1999) Green et al. (2006) Green et al. (2007) Zeltyn et al. (2009) Green et al. (2011)	Blair et al. (1981) Ridge et al. (1998) Kim et al. (1999) Smith et al. (2008) Cochran et al. (2009) Mandelbaum et al. (2010)
ICU	Mcmanus et al. (2004) Griffths et al. (2006) Cochran et al. (2008) Bretthauer et al. (2011) Seshaiah et al. (2011)		Maartje et al. (1999) Chan et al. (2011)
O & N	Milliken et al. (1972) Kao et al. (1981) Cochran et al. (2006)		Asaduzzaman (2010)
G & M	Gorunescul et al.(2002) Koizumi et al. (2005) Chaussalet et al.(2006) Young et al. (1962)	Gupta et al.(1971)	
Other Dept.	Green et al. (2001) Xiaodong Li (2009) Osorio et al. (2009) Bruin et al. (2010)	Khan et al. (1993) Vericourt et al. (2011)	Broyles et al. (2008)

ED-emergency deptartment, ICU-intensive care unit , O & N -obstetrics and neonatal unit, G & M-geriatric and mental health care unit

Bed management

The demand for hospital or clinic beds can be decomposed into both routine (scheduled) and emergency (unscheduled) admissions. Both these types of admissions impact how many beds are needed to meet demand, while maintaining reasonable bed utilisation rates. In the literature, most bed planning queueing models attempt to overcome bed shortages or policies that lead to patient misplacement, bumping, or rejection. Hospitals are faced with the trade off between having available beds to service patient demand vs keeping bed occupancy (utilisation) rates high. Bruin et al. (2005) apply a stationary 2-D queueing system with blocking to analyse congestion in emergency care chains. Their primary goal is to determine the optimal bed allocation over the emergency care chain, given a required service level (max. 5 % refused admissions). They successfully identify bottlenecks, describe the impact of fluctuation in demand and calculate the optimal bed capacity distribution. Cooper et al. (1974) deal with the same problem extended to a sequence of two stations each of which should have a maximum turn-away rate of 5%. Cochran et al. (2006) propose a multi-stage stochastic methodology to balance inpatient bed unit utilisations in an entire hospital. It minimises blocking of beds from upstream units, within given constraints on bed reallocation, while considering multiple patient types. Queueing network analysis and optimisation are used to achieve balanced targets of bed unit utilisation while building hospital staff involvement. Discrete event simulation is then used to maximise flow through the system including nonhomogeneous effects of daily and hourly peak loading, nonexponential lengths of stay, and blocking behaviour. A 400 plus bed major hospital is analysed with the methodology and results are validated against field data. Bruin et al. (2007) investigates the bottlenecks in the emergency care chain of cardiac in-patient flow. Their primary goal is to determine the optimal bed allocation over the care chain given a maximum number of refused admissions. Linda Green (2002) examines data from New York state and uses queueing analysis to estimate bed unavailability in intensive care units (ICU) and obstetrics units. Using various patient delay standards, units that appear to have insufficient capacity are identified. The results indicate that as many as 40% of all obstetrics units and 90% of ICUs have insufficient capacity to provide an appropriate bed when needed. This contrasts sharply with the expectations deduced acording to standard average ocupancy targets. Shmueli et al. (2003) propose a model to maximise the expected incremental number of lives saved by operating an ICU. They use single-queue models to find the probability distribution of the number of occupied ICU beds. They modeled the ICU at Jerusalems Hebrew University-Hadassah Hospital by using the proposed methodology and showed that a relative life saving improvement of $17.9\ \%$ could be achieved by reforming the ICU admission policy. McManus et al. (2004) collect 2 years admission, discharge, and turn-away data in a busy, urban ICU. Using queueing theory, they construct a mathematical model of patient flow, compared predictions from the model to observed performance of the unit, and explored the sensitivity of the model to changes in unit size. Griffths et al. (2006) propose a $M/H/c/\infty$ queueing model of the ICU environment, with particular emphasis on adequately representing the high variation in the patient length of stay. It is anticipated that the model will be utilised as a tool for resource management. Cochran et al. (2009) statistically examine four distinct hospital inpatient data sets for internal consistency and potential usefulness for estimating true patient bed demand. They conclude that posterior financial data, billing data, rather than the census data commonly relied upon, yields true hospital bed demand. Seshaiah et al. (2011) develops a queueing network model with blocking and reneging to study how the wait times in Emergency care units are influenced by the number of available beds in ICU and general unit. They study the $M/G/k$ queue with reneging and obtained a relationship the percentage of reneged patients and the reneging parameter in addition to finding the wait time distribution. Through an approximate logical methods and simulation, they determined the adequate bed counts in each of the two units so as to guarantee certain access standards $M/G/k$. Milliken et al. (1972) seek a 1% turn-away rate in an obstetrics department in which vaginal births have priority over scheduled caesarian sections. They point out the benefits of economies of scale so that larger facilities incur lower bed investment per additional birth. Kao et al. (1981) propose an approach based on queueing models to periodically reallocate beds to services to minimise the expected overflows. They show that queueing based models are simple and useful tools for analysing the bed allocation problem in a health care facility. Cochran et al. (2006a) analyse a 400 plus bed majror hospital, using a queueing network model and optimisation to balance inpatient bed unit utilisations in a hospital. They aim to balance bed unit utilisations across an obstetrics hospital and minimise the blocking of beds from upstream units within given constraints on bed reallocation. Gorunescul et al. (2002) use a queueing model to help plan bed allocation in a department of geriatric medicine. They built an $M/M/c/K$ model to demonstrate how changing admission rates, length of stay, and bed allocation influence bed occupancy of the unit. Their results show that $10-15\%$ bed emptiness is necessary to maintain service efficiency. Koizumi et al. (2005) use a blocking methodology to model patient blocking in a mental health system, To block the resources from patient use, they implement a modification of the service time. They also use simulation to test the robustness of their model. It represents the first application of queueing with blocking to the analysis of congestion in a mental health system. Chaussalet et al. (2006) apply a closed queueing network to modeled patient flow in health care systems with bed capacity constraints in order to provide a useful decision aid for health service managers. Bailey (1954) uses a queueing model to find the number of beds and servers in a hospital. Young (1962) proposes an incremental analysis approach in which the cost of an additional bed is compared with the benefits it generates. Beds are added until the increased cost equals the benefits. Linda Green et al. (2001) apply a queueing model approach to the hospital bed planning issue to gain insights on the potential impact of cost-cutting strategies on patients' delays for beds. Using this approach, they also identify those factors that have the greatest impact on the trade-off between hospital occupancy levels and delays. De Bruin et al. (2010) develop a decision support system, based on the Erlang loss model, which can be used to evaluate the current size of nursing units and to quantify the impact of bed reallocations and merging of wards. Xiaodong Li et al. (2009) introduce a decision aiding model for optimising the allocation of beds in a hospital based on queueing theory and goal programming. The $M/PH/m$ queueing model can be used effectively to investigate some essential characteristics of access to service in the hospital. It provides a means of calculating patient admission probability and profit achievement as functions of the number of beds for each individual department. Carolina Osorio et al. (2009) present an analytic network model that preserve the finite capacity of real system. The model is formulated for multiple server finite capacity queueing networks with an arbitrary topology and blocking-after-service. It is validated by comparison with both pre-existing methods and simulation results. It is then applied to study patient flow in a network of units of the Geneva University Hospital. The model allows for an identification of three main sources of bed blocking and to quantify their impact upon the different hospital units. Jones et al. (2011) investigates the adequacy of current models used to forecast bed demand and explore the issues surrounding the correct level of occupancy required to deliver effective and safe health care.

Staff management

Staff planning in order to satisfy the demand is one of the areas in which queueing models can be used effectively. Queueing models can help planners to estimate the number of required staff in each unit to achieve an acceptable customer LWTR. Cheang et al. (2003) review nurse rostering problems, while Ernst et al. (2004) review the staff scheduling and rostering.

Agnihothri et al. (1991) utilise a $M/M/C$ queueing model to find the optimal staffing levels to handle the variation in call arrivals to an appointment system. They found that the existing staff and the number of hours they were working was enough to handle the demand and by redistributing server capacities over time, they could effectively reduce customer complaints. In general, the nurse scheduling problem is that of assigning shifts to nurses having different skills while satisfying as many soft constraints and personal preferences as possible. A nurse schedule will typically strive to meet the required personnel coverage over a predefined planning period. Linda Green et al. (2006) use a queueing analysis to identify provider-staffing patterns in order to minimise leaving without treatment ratio (LWTR). They showe that despite an increase of 6.3% in patient arrivals, through a weekly 3.1% increase in staffing, LWTR could be decreased by 22.9%. Linda Green et al. (2007) review queueing-theory methods for setting staffing requirements in service systems where customer demand varies in a predictable pattern over the day. Analysing these systems is not straightforward, because standard queueing theory focuses on the long-run steady-state behavior of stationary models. They show how to adapt stationary queueing models for use in non stationary environments so that time-dependent performance is captured and staffing requirements can be set. Zeltyn et al. (2009) evaluate ED staff scheduling that adjusts for mid-term changes (tactical horizon, several weeks or months ahead). Also they analyse the design and staffing problems that arose from physical relocation of the ED (strategic yearly horizon). Natalia Yankovic et al. (2011) present a two-dimensional queueing model to guide nurse staffing decisions and demonstrating its reliability in identifying good staffing levels across a broad range of parameters corresponding to actual hospital units. The model is also useful for estimating the impact of nurse staffing levels on emergency department overcrowding. Tucker et al. (1999) consider activating a second operating room (OR) team during the night shift. Using queueing theory, they find that the probability of two patients needing the operating room services is negligible. Gupta et al. (1971) illustrates the practical application of queueing theory to a simple problem in manpower planning: how large a staff is required to give adequate service from a hospital messenger unit? The problem is simple enough to be amenable to an analytic solution, and the optimal solution arrived at by the use of queueing theory resulted in a considerable saving to the hospital studied. Riaz Khan et al. (1993) incorporate advertising into their model to control the demand for laboratory services. For each staffing level, they determine the number of clients that would maximise profits. They then choose the staffing level with maximum profits and apply the necessary amount of advertising that would attract the desired number of clients. The model assumes that clients would leave without service if they wait above a certain amount of time. De Vericourt et al. (2011) present a closed queueing model to determine efficient nurse staffing policies. Their approach is to explicitly model the workload by s nurses within a single medical unit with n homogeneous patients as a closed $M/M/s//n$ queueing system and deduce efficient nurse staffing rules using many server asymptotic approach. Their objective is to determine the number of nurses that should be present at any time in the medical unit. Bretthauer et al. (2011) consider the problem of optimal capacity allocation in a hospital setting, where patients pass through a set of units, for example intensive care and acute care (AC), or AC and post-acute care. If the second stage is full, a patient whose service at the first stage is complete is blocked and cannot leave the first stage. They develop a new heuristic for tandem systems to efficiently evaluate the effects of such blocking on system performance and they demonstrate that these heuristic performs well when compared with exact solutions and other approaches presented in the literature. In addition, they show how their tandem heuristic can be used as a building block to model more complex multi-stage hospital systems with arbitrary patient routing, and they derive insights and actionable capacity strategies for a real hospital system where such blocking occurs between units.

Room facility management

The continuing trend towards the development of free standing surgicenters, as well as the movement to deliver health care services away from inpatient facilities and more towards outpatient facilities, has put increased pressure upon hospital management to expand their outpatient services and/or to build new facilities to handle these additional patients.

Blair et al. (1981) use a finite capacity multi-server queues and continuous time Markov chains to model a system of burn care facilities linked together by a referral policy to accommodate patient overflow. They utilise a heuristic optimisation procedure to find the optimal setting in a burn care facility in New York State, where the goal is to maintain a 95 % level of service and keep LWTR less than 5 %. They found that their proposed approach is ideal for a system with low demand and high infrastructural costs. Mayhew et al. (2008) use a queueing model to evaluate the emergency departments of hospitals in United Kingdom. Per government mandated targets, they focused on completing and discharging 98% of patients within (4hrs) four hours. Avishai Mandelbaum et al. (2010) study routing algorithms that are applicable to assigning hospital patients, from the emergency department to internal wards. Cochran et al. (2009) use an open queueing network model to design an emergency department in order to increase its capacity. They use waiting time and over flow probability as system performance indicators in their work. Also they conclude that population growth, unavailability of emergency departments nearby and variation of seasonal peak can considerably increase patients average waiting time, which leads to an increase in LWT. Broyles et al. (2008) present a queueing network methodology to aid decision makers in capacitating their outpatient clinic given variable patient demand and inconsistent patient mix. The methodology uses a general service distribution, open queueing network model to estimate clinic room utilisation and expected patient wait times. Ridge et al. (1998) use an $M/M/c$ model (with priority to emergency patients over elective patients) to analyse a 6 bed intensive care unit (ICU). They calculate the waiting time in queue for both the emergency patients and the elective patients. The queueing model is used for the purposes of validating a simulation model. Kim et al. (1999) use an $M/M/c$ queueing model to analyse the capacity of a 14 bed ICU. They had four types of patients in their model for which they evaluate three different ways of computing the overall average service time for the queueing model. However they exclude patients with a very long stay, and the transfer of patients in their model. The results provide insights into the operations management issues of an ICU facility to help improve both the unit's capacity utilisation and the quality of care provided to its patients. Maartje et al. (2009) propose a queueing model to reschedule the appointments and the reallocate the redesign the hospital preanesthesia evaluation clinic. The intervention resulted in a shortening of the time the anesthesiologist needed to decide upon approving the patient for surgery. Patient arrivals increased sharply over 1 yr by more than 16%; however, patient length of stay at the clinic remained essentially unchanged. If the initial set-up of the clinic had been maintained, the patient length of stay would have increased dramatically. Carri W. Chan et al. (2011) examine the queueing dynamics of an ICU where patients may be readmitted. When patient demand exceeds availability, current ICU patients may be discharged in order to accommodate new, more urgent patients. Such a discharge increases the likelihood of readmission to the ICU. They modeled such an ICU as a state-dependent Erlang-R queueing network where service times and readmission probabilities depend on whether the ICU is full. They consider the definition of full affects system behavior and provide insight into capacity management of such systems. Md Asaduzzaman et al. (2010) propose a queueing model to determine the number of cots at all care units for any desired overflow and rejection probability in a neonatal unit. They derive a solution to the capacity problem faced by many perinatal networks in the United Kingdom.

4.2.2 Patient scheduling

Patient scheduling and admissions focus on procedures that determine how patient appointments (with medical staff) are scheduled, both in terms of when and how they are set in a given day, and their length of time. More specifically, this involves rules that determine when appointments can be made (namely, morning vs afternoon) and the length (spacing) of time between appointments. This may also be extended to include designating the specific type of medical staff who will be responsible for treating patients and the clinic space that will be required to deliver the necessary treatments.

Appointment systems

The appointment system is usually applied when admitting patients to hospital wards or registering persons in outpatient clinics. The overall problem is to control patients waiting times by applying an appropriate, sensitive, and responsive appointment procedure in which the estimated patient flow corresponds to the available resources. Some authors try to minimise the delays after patients arrive in the clinic, while others concentrate on the time that patients have to wait to get an appointment. Patient appointment systems are highly correlated with patients waiting time and server utilisation. Since decreasing patients waiting time and increasing expensive servers utilisation are primary goals of a health care provider, it is necessary for them to design and implement an efficient appointment system to be able to improve the quality of their service.

Bailey (1952, 1954) proposes (a) appointment interval and (b) consultant arrival time as two variables that determine the efficiency of an appointment system. In order to find a balance between patient wait time and consultant idle time, the first step is to determine the relative values of patient time and consultant time. The ratio of the total time wasted by all patients to the consultant idle time should equal the value of the consultants time relative to the patients. He chooses to assign individual appointment times at intervals equal to the average patient processing time and finds that the consultant should arrive at the same time as the second patient. Albin et al. (1990) show how a queueing network model helped to uncover the causes of delay in a health center appointment clinic. Brahimi et al.(1991) design an appointment system to reduce the number of patients in the queue at any time, and reduce patient waiting time without significantly increasing doctor idle time. De Laurentis et al. (2006) use a queueing network and simulation to study an open access appointment scheduling system at an urban outpatient clinic. They consider a same-day appointment system and showed that the pre-scheduling horizon and the percentage of patients using open access scheduling are key factors for a successful open access scheduling policy. They also suggest that clinics with many visiting doctors, such as residents, are not good candidates for the same-day open access appointment system. In Vasanawala et al. (2005) a radiology department has some time slots scheduled for routine radiology analysis. Emergency requests may require rescheduling of scheduled requests. Given a 1 % or 5% probability of rescheduling, they used queueing theory to determine how many scheduled slots to leave empty during routine scheduling. Linda Green (2010) conceptualise an appointment system as a single-server queueing system in which customers who are about to enter service have a state dependent probability of not being served and may rejoin the queue. Linda Green derives stationary distributions of the queue size, assuming both deterministic as well as exponential service times, and compare the performance metrics to the results of a simulation of the appointment system. Bruin et al. (2010) apply a approximations based on the infinite-server queue, they analyse an $M_t/H/s/s$ model to determine the impact of the time-dependent arrival pattern on the required number of operational beds and fraction of refused admissions for clinical wards. Their main goal is to analyse the impact of a predictable patient arrival pattern on the performance and bed capacity requirements of a clinical ward. Rene Bekker et al. (2011) analyse the impact of the sources of variability on the required amount of capacity and to determine admission quota for scheduled admissions to regulate the occupancy pattern. For the impact of variability on the required number of beds, they use a heavy-traffic limit theorem for the $G/G/$ queue yielding an intuitively appealing approximation in case the arrival process is not Poisson. Also, given a structural weekly admission pattern, they apply a time dependent analysis to determine the mean offered load per day. Jackson et al. (1964), apply the queueing system to the problem of out-patient and general practice appointment systems. The fundamental problem is the case where patients are given appointments to see one or more doctors. The patients may arrive on time, early, late, or not at all and are seen by a doctor in order of arrival. The two primary performance parameters of the system are patients' waiting time and doctor's idle time. They tackle this problem by showing quite clearly how in practice appointment systems are often designed massively in favour of reducing doctor's idle time at the expense of increasing the patients' waiting time. Also, they produce results using queueing theory and using simulation

4.3 Health care system analysis

Queueing models are used to gain information about activities within a health care system (waiting times, utilisation rates, queue times and lengths) and to indicate the reasons for mistakes in patient care and during the treatment processes. They may help to forecast the consequences of decisions made, to detect unfavorable trends, and to make improvements to the general performance of a system. Models from this application area are usually directed at hospital administrators and perform the role of an information system. We classified these into three sub areas like waiting time and utilisation analysis, minimisation of cost, queue times and lengths.

4.3.1 Waiting time and utilisation analysis

In a queueing system, minimising the time that customers (in health care, patients) have to wait and maximising the utilisation of the servers or resources (in health care, doctors, nurses, hospital beds, e.g.) are conflicting goals. We are dividing them into four categories: Leaving without treatment ratio (LWTR), variable arrival rate (VAR), priority queue discipline (PQD), minimum waiting time (MWT).

Leaving without treatment ratio (LWTR)/Reneging

Solberg et al. (2003) listed 38 measures for evaluating emergency department performance. One the most important measures mentioned in this work is leaving without treatment ratio (LWTR), which has been studied by several other researchers. When a patient is waiting in a queue, the patients may decide to forgo the service because the patients does not wish to wait any longer. This phenomenon, called reneging, is an important characteristic of many health care systems. The probability that a patient reneges usually increases with the queue length and the patient's estimate of how long he must wait to be served. In systems where demand exceeds server capacity, reneging is the only way that a system attains a state of dysfunctional equilibrium (see Hall et al. (2006). An important example of such a system is an emergency department

Table 3: Classification scheme: health care system design and system analysis

System analysis	Emergency Departments	Other Departments	Pharmacy
LWTR	Worthington et al. (1987) Solberg et al. (2003) Green et al. (2006) Broyles et al. (2007) Roche et al. (2007) Cochran et al. (2009) Cochran et al. (2010)	Collings et al. (1976) Zenios et al.(1999) Seshaiah et al. (2011)	
VAR	Taylor et al. (1969) Haussmann et al.(1970) McQuarrie et al. (1983) Panayiotopoulos et al. (1984) Rosenquist et al. (1987) Worthington et al. (1991) Adeleke et al. (2009) Siddhartan et al. (1996)	Milliken et al. (1972)	Shimshak et al. (1981) Ndukwe et al. (2011)
PQD	Au.Yeung et al. (2007) Fiems et al. (2007) Laskowki et al. (2009) Mokadis et al. (2011) Creemers et al. (2011)	Moore et al. (1977) Weiss et al. (1987) Huang et al. (1994)	Vemuri et al. (1984) Moss et al. (1987)
MWT		Jiang et al.(2008) Goddard et al. (2008) Igor et al. (2009) Joustra et al. (2010)	Lin et al. (1996) Boyce et al. (1998) Lin et al. (1999)

Leaving without treatment ratio (LWTR), variable arrival rate (VAR), priority queue discipline (PQD), minimum waiting time (MWT).

Broyles et al. (2007) calculate the percentage of patients who leave an emergency department without getting help using arrival rate, service rate, utilisation, capacity. From this percentage, they determine the resulting revenue loss. It is possible to redesign a queueing system to reduce reneging. A common approach is to separate patients by the type of service required. Roche et al. (2007) find that the number of patients who leave an emergency department without being served is reduced by separating non-acute patients and treating them in dedicated fast-track areas. Most of their waiting would be for tests or test results after having first seen a doctor. The paper also estimates the size of the waiting area for patients and those accompanying them. Cochran et al. (2010) propose to enable strategic decision making on future ED capacity on the basis of patient safety (rather than congestion measures). They hypothesise that the LWTR reneging percentage is captured by the balking probability relationship of an $M/M/1/K$ queue. If true, this relationship is superior to the typical ad hoc regression relationships commonly found. Zenios et al. (1999) develop a queueing model with reneging that provides a stylistic representation of the transplant waiting list. The model assumes that there are several classes of patients, several classes of organs, and patient reneging due to death. They focus on randomised organ allocation policies and develop closed-form asymptotic expressions for the stationary waiting time, stationary waiting time until transplantation, and the fraction of patients who receive transplantation for each patient class.

Variable arrival rate (VAR)

Although most analytical queueing models assume a constant customer arrival rate, many health care systems have a variable arrival rate. In some cases, the arrival rate may depend upon time but be independent of the system state. For instance, arrival rates change due to the time of day, the day of the week, or the season of the year. In other cases, the arrival rate depends upon the state of the system.

Collings et al. (1976) discuss the queueing problem of an intensive care unit and show that it is unlikely that the hourly variation in the arrival rate of patients to the unit will significantly affect the number of beds occupied. A system with congestion discourages arrivals. Worthington (1987) presents an $M(q)/G/S$ model for service times of any fixed probability distribution and for arrival rates that decrease linearly with the queue length and the expected waiting time. The arrival rate may increase over time due to population growth or other factors. Rosenquist (1987) studies the characteristics of examination time and interarrival time that are determined for an emergency room radiology service, analysing two examples that demonstrate the usefulness of queueing analysis in radiology. In the first example the effects of a 5% annual increase in patients on waiting time are shown, while the second example shows the use of the technique for cost analysis. Radiologists and medical administrators should be aware of the availability and value of queueing analysis for department planning and management. Worthington (1991) suggests that increasing service capacity (the traditional method of attempting to reduce long queues) has little effect on queue length because as soon as patients realize that waiting times would reduce, the arrival rate increases, which increases the queue again. Adeleke (2009) considers the waiting of patients in university health centers as a single-channel queueing system with Poisson arrivals and exponential service rate where arrivals are handled on a first come first serve basis. Using $M/M/1$ queueing system, they obtained the average number of patients and the average time spent by each patient as well as the probability of arrival of patient into the system.

Priority queueing discipline (PQD)

In most health care settings, unless an appointment system is in place, the queue discipline is either first-in-first-out or a set of patient classes that have different priorities (as in an emergency department, which treats patients with life-threatening injuries before others). Queueing discipline is an important factor that may significantly affect waiting time for service.

Taylor et al. (1969) model an emergency anaesthetic department operating with priority queueing discipline. They are interested in the probability that a patient would have to wait more than a certain amount of time to be served. Haussmann (1970) investigates the relationship between the composition of prioritised queues and the number of nurses responding to inpatient demands. The research finds that slight increases in the number of patients assigned to a nurse and/or a patient mix with more high-priority demands result in very large waiting times for low priority patients. Taylor et al.(1980) consider a priority queue in steady state with N servers, two classes of customers, and a cutoff service discipline. Low priority arrivals are "cut off" (refused immediate service) and placed in a queue whenever N_1 or more servers are busy, in order to keep $N-N_1$ servers free for high priority arrivals. A Poisson arrival process for each class, and a common exponential service rate, are assumed. Two models are considered: one where high priority customers queue for service and one where they are lost if all servers are busy at an arrival epoch. Results are obtained for the probability of n servers busy, the expected low priority waiting time, and (in the case where high priority customers do not queue) the complete low priority waiting time distribution. The results are applied to determine the number of ambulances required in an urban fleet which serves both emergency calls and low priority patient transfers. McQuarrie (1983) shows that it is possible, when utilisation is high, to minimise waiting times by giving priority to clients who require shorter service times. This rule is a form of the shortest processing time rule that is known to minimise waiting times. It is found infrequently in practice due to the perceived unfairness (unless that class of customers is given a dedicated server, as in supermarket check-out systems) and the difficulty of estimating service times accurately. Panayiotopoulos et al. (1984) develope a methodology based on $GI/G/C(t)$ model to find the adequate number of servers for reducing waiting time in an emergency department. They also consider the limited waiting room, patients priorities and single visit of the system by each server within a certain period of time in the model. Siddhartan et al. (1996) investigates the increased waiting time costs imposed on society due to inappropriate use of the emergency department by patients seeking non-emergency or primary care. They propose a priority discipline for different categories of patients and then a first-in-first-out discipline for each category. They find that the priority discipline reduces the average wait time for all patients: however, while the wait time for higher priority patients is reduced, lower priority patients endure a longer average waiting time. Au-Yeung et al. (2007a) develop a multi-class Markovian queueing network model of patient flow in the Accident and Emergency department of a major London hospital. Using real patient timing data to help parameterize the model, they solve for moments and probability density functions of patient response time using discrete event simulation. Fiems et al. (2007) investigate the effect of emergency requests on the waiting times of scheduled patients with deterministic processing times. It is a pre-emptive repeat priority queueing system in which the emergency patients interrupt the scheduled patients and the latter's service is restarted as opposed to being resumed. Marek Laskowski et al. (2009) apply both agent-based models and queueing models to investigate patient access and patient flow through emergency departments.The models are developed independently, with a view to compare their suitability to emergency department simulation. The framework of multiple-priority queue systems and the genetic programming paradigm of evolutionary machine learning are applied as a means of forecasting patient wait times and as a means of evolving health care policy, respectively. Mokadis et al. (2011) develop a multi-class Markovian queueing network model of patient flow in the Accident and Emergency department of a major in the Egypt Health Service. Using a discrete event simulation, they investigate the impact of giving priority treatment to different classes of patients, and compare the resulting response-time densities and moments with real data. Creemers Stefan et al. (2011) discuss the difference between pre-emptive and non-pre-emptive outages. They concentrate on service outages and develop new expressions to assess their impact on waiting lists and delays. Using data obtained from a Belgian hospital, the expressions are evaluated through a number of queueing models.

Minimum waiting time (MWT)

Moore (1977) reduces customer waiting time for birth and death certificates at the Dallas bureau of vital statistics by decreasing the time required to serve each customer. This research first uses queueing theory to calculate the service rate required to achieve a target waiting time of 15 minutes. Weiss et al. (1987) use a queueing-analytic approach to describe the process by which patients await placement. They model the situation using a state-dependent placement rate for patients backed up in the acute care facility. They compare their model results with data collected from a convenience sample of 7 hospitals in New York State. Huang et al. (1994) report the results of a survey on patient attitude towards waiting in an outpatient surgery clinic. Jiang et al. (2008) analyses the impact of parallelisation on cycle time of patients in the health care system using multi-class open queueing network model. Goddard et al. (2008) develop a queueing model of public sector hospital waiting lists. The model's equilibrium solution is sensitive to the criteria determining which patients gain access to the waiting list, and the order in which those patients are admitted to hospital. Igor Georgievskiy et al. (2009) examine the admissions process in a regional hospital with the purpose of documenting the existing process and its bottleneck points, determining the waiting time distributions at the two waiting room settings within the system, and developing recommendations for modifying the layout and staffing of the system to reduce waiting times for patients. Joustra et al. (2010) examine whether urgent and regular patients waiting for a consultation at a radiotherapy outpatient department should be pooled or not. The practical approach indicates that the separation of queues may require less capacity to meet the waiting time performance target for urgent as well as regular patients.

4.3.2 Queuelength / Limited Queue discipline (LQD)/ Blocking

Some studies consider the situation where the queue length is limited. When the queue reaches the maximum length allowed, walk-in patients will be turned away until the queue length decreases. This event is called blocking.

McManus et al.(2004) construct a mathematical model of patient flow in a busy urban intensive care unit. The proposed queueing model could predict admission turn-away (blocking) accurately with correlation coefficient of 0.89. Using the proposed queueing model, they also find that the system performance would drastically decrease with even a small change in servers availability. Koizumi et al. (2005) apply a queueing network system with blocking to analyse the health care system congestion processes. Both mathematical and simulation results are presented and compared. They found that one of the main reasons for system congestion is existence of facility-specific bottlenecks and removing such bottlenecks may efficiently reduce congestion in the system. Kapadia (1985) presents the queueing processes of waiting lines with two priorities and multiple service channels. The arrival process is assumed Poisson and the service time distribution is negative exponential. Arriving units enter service if there is at least one idle channel, otherwise they join a finite queue and are served according to a non pre-emptive priority discipline. If a low priority arriving unit finds the queue full, it is not allowed to enter the system and is considered blocked or lost. In the first model a high priority arrival may displace a low priority unit from the full queue and the latter may be blocked if the queue consists of high priority units only. In the second model the high priority unit may still displace a low priority unit from the full queue but it will never be blocked and may wait outside the system if the system is full. Thus far there has been no discussion of such models in queueing theory literature. The analytical expressions for average waiting times have been obtained for the two models and two potential applications of the models are described and the usefulness of the models is illustrated by numerical examples.

4.3.3 Minimise Costs

Most of the research assigns costs to patient waiting time and to each server. After modelling the system using queueing theory, minimizing costs reduces to an exercise of finding the resource allocation that costs the least or generates the most profit.

Young (1962a, 1962b) proposes an incremental analysis approach in which the cost of an additional bed is compared with the benefits it generates. Beds are added until the increased cost equals the benefits. Keller et al. (1973) set out to determine the capacity with minimal costs required to serve patients at the Duke University Medical center. They find that the current capacity is good but needs to be redistributed in time to accommodate patient arrival patterns. Gorunescu et al. (2002) use backup beds (only staffed during peak demand) to reduce the probability of patient turn-away at a marginal cost. The model assumes a phase-type service distribution. Paula Gonzalez et al. (2004) deals with a cost-allocation problem arising from sharing a medical service in the presence of queues. They use a standard queueing theory model in a context of several medical procedures, a certain demand for treatment and a maximum average waiting-time guaranteed by the government. They show that the sharing of an operating-theatre to treat patients of different medical disciplines leads to a cost reduction. John. K. Obamiro (2010) presents the general background of queueing theory and its associated terminology, and how queueing theory can be used to model an ante-natal clinic of a Federal teaching hospital located in Lagos, Nigeria. The resultant performance variables can be used by the policy makers to increase efficiency, improve the quality of care, as well as decrease cost in hospital organisations and services.

5 Simulation-Based queueing models in health care

Queueing models and simulation models each have their advantages. Queue modelling is potentially very valuable, as many real situations in important industries can be formulated as queueing/service systems. Over the last 50 years or so in the Operational Research Quarterly and the Journal of the Operational Research Society the most frequently reported application area have been health care. In health care, for example Bailey (1957), when writing generally about hospital planning and design, highlighted the scope of operational research methods to extract the maximum amount of benefit for the community out of restricted resources, citing appointment systems and emergency beds as examples.

Simulation modelling has long been part of queue modelling, and many of the applications and some of the research described earlier in this paper have made some use of simulation models. Drawing a parallel with analytic formulations and numerical solutions, simulation models are also capable of providing performance measures for a particular queueing system with specific parameter values, and hence can similarly also be viewed like formulae An added caveat is that the results will have confidence intervals, although these can be very narrow by having sufficient runs of the model. Unlike numerical methods, simulation cannot rely on underlying theory to provide understanding and insights into the behaviour of the queueing systems being modelled, but instead needs to rely on the weight of empirical evidence. This, however, can be very effective, especially in a practical situation where managers might react better to empirical evidence than to mathematically derived insights. For example, in a wholly simulation based study of queueing problems in UK accident and emergency (A & E) departments, Fletcher et al (2007) achieved general recognition that the NHS target requiring 98% of patients to complete their stay in A & E within 4 hours was reasonable, and general understanding of what needed to be done in individual departments for that to be possible, such as matching staffing levels to underlying arrival rates.

In a research setting the flexibility of simulation modelling is a major asset; it can provide a way of investigating the accuracy of many analytic formulations and solutions that have simplified a queueing problem, in some respect, to make the mathematical analysis feasible. In a practical setting, the flexibility is also potentially valuable in its ability to cater for particular features of the problem of interest. Hence, operational researchers have attempted to combine simulation with queueing techniques, to capitalise on the advantages of using both techniques simultaneously. Also a combination of queueing theory and simulation let to practical insights and results, and seem highly fruitful in health care system. The advantage of hybrid queueing/simulation models is discussed in van Dijk (2000) and van Dijk. et al. (2008). Jun et al. (1999) complete a survey of discrete event simulation models in health care citing over 100 articles and discussing the various applications in clinical settings. It is also interesting to note that while simulation modelling has only become easily available and user friendly in recent years, Bailey (1952), in one of the earliest studies of hospital outpatient clinics, phrased the problem in the language of queueing theory but chose to tackle it by using a pre-computer simulation to model the consequences of different appointment systems. Kao and Tung (1981) and Tucker et al. (1999) use simulation to validate, refine or otherwise complement the results obtained by queueing theory. Albin et al. (1990) show how one can use queueing theory to get approximate results and then use simulation models to refine them. In the last decade the number of health care problems that have been studied using a queueing network approach has increased tremendously.

Table 4: Categorization of references

Method	Authors
Complex queueing network model with simulation	Albin et al. (1990), Koizumi et al. (2005), Aaby et al. (2006),Cochran et al. (2006a), De Laurentis et al. (2006), Au-Yeung (2007), Creemers et al. (2007), Jiang et al. (2007), Ali Pilehvar et al.(2008), Cochran et al. (2009), Osorio et al. (2009), Seshaiah et al. (2011)
Simple queueing models and its applications with simulation	Ladany (1978), Kao and Tung (1981), Vemuri (1984), Babes et al. (1991, Ho et al. (1992), Liyanage et al. (1995), Lin et al. (1996), El-Darzi (1998) Ridge et al. (1998), Zenios et al. (2000), Shmueli et al. (2003) Green (2003), Griffiths et al.(2004), Green (2005), Abellan et al. (2006), Green et al. (2008), Wang et al. (2008), Igor Georgia et al. (2009), Karnon (2009), Zeltyn et al. (2009), Laskowski (2009), Zai et al. (2009), Eugene Day et al. (2010) Joustra et al. (2010), Palvannan et al. (2010), De Vericourt et al. (2011), Yankovic et al. (2011)

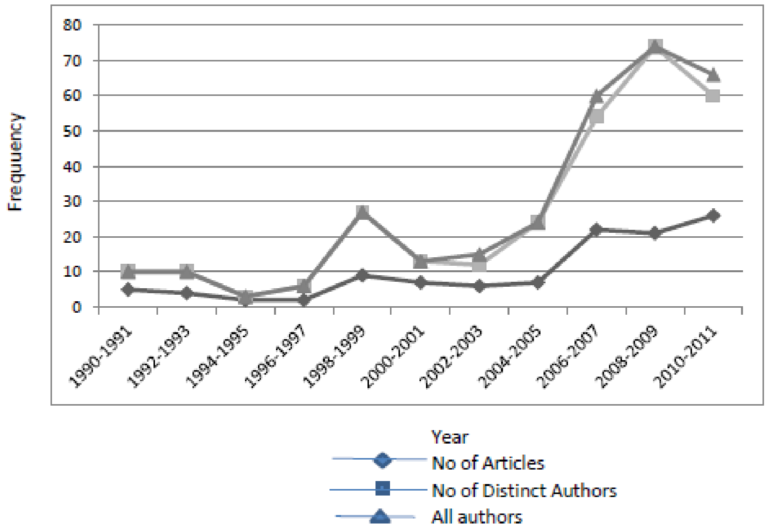
6 Descriptive analysis

In general, the nature of the data available in the studies reviewed determines the types of meta-analytic methods that can be applied. In this paper, the simple meta-analysis provides only descriptive information with no statistics, it is expected to shed greater understanding on the development and evolution of QTA research in HCM and to identify potential research areas for further research and for improvement. We identified and analysed the 141 articles related to QTA research in HCM by (1) year of publication,(2) articles in the journals (discipline wise), and (3) using our classification groups.

6.1 Distribution of articles by the year of publication

We identified 31 articles published between 1952-1989 and the remaining 110 articles are from 1990 to 2011. The distribution of articles published with the class interval of two years is shown in Fig. 2 from 1990 to 2011. The class interval of two years is selected because any article takes about 1 to 2 years on the average to publish in the reputed/international journals. Figure 2 indicates that there is limited research outputs before 2000. Effectively there is an increasing research trend on queueing theory application in health care fom 2000 onward. This increased research trend on QTA research in HCM along with the number of researchers contributed is also shown in Fig. 2. Another possible reason for the continuous increased trend from 2000 is that the advancement in computational power and software availability.

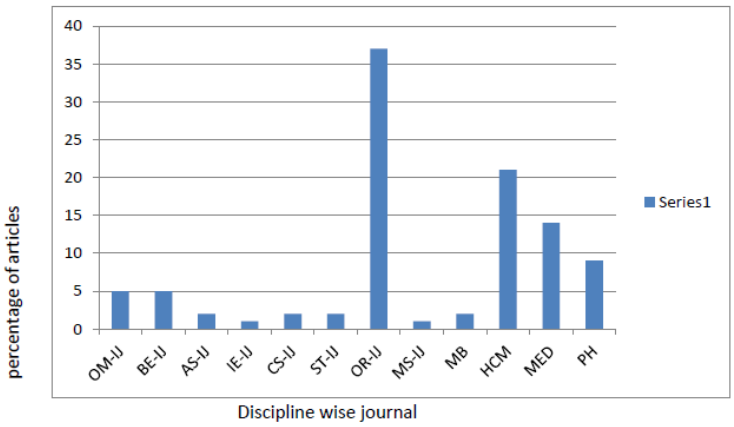
Fig 2. Total number of articles, distinct authors and all authors by annually



6.2 Distribution of articles by journals (discipline wise)

In journals and other possible publication outlets, the number of QTA research articles in HCM reported is computed and shown in Table 1. During the period 1952-2011 there are 141 research articles on QTA research in HCM which appeared in 61 journals and other publication outlets. Of these, 112 articles are from journals, 17 articles are from various proceedings of the conference, two from working papers and the submitted to journal category, one from lecture notes, 8 from technical reports, one from book. Since the frequency of the pertinent articles cross many journals is low, the research journals reviewed are grouped with respect to disciplines: viz., International Journal (IJ) on Operations Research (OR), Industrial Engineering (IE), Computer science and Simulation (CS), OM-Operation Management, BE-Business and Economics, AS- Applied statistics, MS-Mathematical Sciences, MSB-Mathematical Biological Sciences, ST-Science and Technology, HCM-Health care Management, MED-Medicine and Ph-Pharmacy. From Fig. 3, it is observed that Operational Research (OR) related journals have by far the most articles. This indicates that the power of OR is continuously explored. Following the OR related journals, the HCM sources are having more articles on queueing theory research in health care management. This could be due to the fact most of the articles reported in OR and HC related journals give little indication on whether the system has been implemented or not. Other possible reasons could be due to the low correlation between "the objective of various studies reported on QTA research in HCM" and the scope of the respective journals. Finally, we observe from both Table 2 and Table 3 that the maximum number of researchers are worked in the field of bed managment as well emergency department in the health care systems.

Fig. 3 Total number of articles (dicipline wise) journals



7 Conclusions

In this paper, applications of queueing theory in modelling hospital processes have been reviewed and categorised. Since health care facilities directly deal with human lives, improving system performance is a very important goal. Increasing servers utilisation and decreasing patients waiting time can enhance system productivity. Queueing theory provides an effective and powerful modelling technique that can help managers achieve the aforementioned goals. This approach can be easily implemented and has several advantages such as providing good and rapid estimations of the system performance.

We presented a review of 141 papers published in peer-reviewed journals and in proceedings from recognized international conferences from 1952 to 2011. Our objective was to identify the leading areas of health care related problems as modelled by queueing theory. This paper reviewed the use of queueing theory for the analysis of different types of health care processes. Models for system design, system operation, estimating waiting time and utilisation and appointment system evaluation were presented. Also, we have reviewed the simulation-based queueing models have been presented. From this review, we found that most QTA articles in HCM were published after 1990. This is because of advancement in computational power and software availability.

Finally, we acknowledge that this review cannot claim to be exhaustive, but it does provide reasonable insights into the state-of-the-art on QTA research in HCM. Thus, we hope that this review will provide a source of reference for other researchers/readers interested in queueing theory applications in health carde management.

Challenges and Directions for Future Research

In the last decade the number of health care problems that have been studied using a queueing network approach has increased tremendously. In this final section we point out a few directions for future research. We distinguish between mathematical challenges- health care problems for which appropriate queueing network models have not yet been developed and health care challenges- health care problems which have not been studied yet, but could be studied with the queueing techniques available in the literature.

The mathematical challenges mainly lie in the modelling aspect. One example is the development of models for networks of care providers who perform several tasks in parallel, in sequence, and sometimes even in a mixed form. Polling models (Takagi 2000) could be of interest here. Also, clinics where patients have to (re-)visit specific care providers in a network of care queues still involve modelling complications. However, re-visiting occurs often in reality (for example, consider the complex network of multiple care providers in ED treatment). The application of time inhomogeneous models that capture the time-dependent arrival patterns of patients has attained only limited attention, see for example (Linda Green et al. 2006). Introducing time in homogeneity in queueing networks is a tremendous challenge. Related is the development of computationally efficient methods that explicitly take into account opening hours of health care facilities. Also, applying simulation-based queueing model is well suited for investigating the spread of infectious diseases, which can reveal the dynamics of an existing or a potential epidemic.

Health care professionals in a couple of fields are familiar with the possibilities of mathematical decision support techniques in general and queueing theory in particular. Our aging population requires increasing care, which has to be delivered with limited resources. Rationing care and the consequences thereof has therefore become an important research topic. Decisions regarding which patient class will be offered what type of care are inevitable. The influence of these decisions on other patient classes, regarding accessibility and other important matters, should be studied in detail. Moreover, the dimensioning of health care facilities, not only in the number of beds required, but also regarding care that will be offered to certain patient classes only, will become increasingly import

References:

[1] Aaby, K., Herrmann, J.W., Jordan, C., Treadwell, M., Wood, K. (2006) Using Operations research to improve mass dispensing and vaccination clinic planning, Interfaces, 36(6), 569-579.

[2] Abell´an, J. J., Armero, C., Conesa, D., P´erez-Panad´es, J., Mart´ınez-Beneito, M. A., Zurriaga, O., Vanaclocha, H., Garc´ıa-Blasco, M. J. (2004) Predicting The Behaviour Of The Renal Transplant Waiting List In The PaÃ-s ValenciÃ (Spain) Using Simulation Modeling, Proceedings Of The 2004 Winter Simulation Conference. R. G. Ingalls, M. D. Rossetti, J. S. Smith, And B. A. Peters, Eds.

[3] Adeleke, R.A., Ogunwale, O.D., Halid O.Y. (2009) An application of queueing theory to waiting time of out-patients in hospitals. Pacific Journal of Science and Technology, 10(2), 270-274.

[4] Agnihothri, S.R., Taylor P.F. (1991) Staffing a centralized appointment scheduling department in Lourdes Hospital. Interfaces, 21, 1-11.

[5] Albin S.L., Barrett, J., Ito, D. and Mueller, J.E. (1990) A Queueing Network Analysis of a Health center, Queueing systems, 7, 51-61.

[6] Md Asaduzzaman, Chaussalet, T.J., Robertson. N.J. (2010), A loss network model with overflow for capacity planning of a neonatal unit, Annals of Operation Research, 178, 67-76.

[7] Au-Yeung, S.W.M ., Harrison, P.G., Knottenbelt, W.J. (2007) A Queueing Network Model of Patient Flow in an Accident and Emergency Department, Valuetools 07, October 23-25, Nantes, France.

[8] Babes, M., Sarma, G.V. (1991), Out Patient Queues At The Ibn-Rochd Health Centre, Journal of Operation Research Society, 42(10), 845-855.

[9] Bailey, N.T.J. (1952) A study of queues and appointment systems in hospital outpatient departments, with special reference to waiting-times. Journal of the Royal Statistical Society, Series B, 14(2), 185-199.

[10] Bailey, N.T.J. (1954) Queueing for medical care, Applied Statistics, 3, 137-145.

[11] Bekker, R., de Bruin A.M. (2010) Time-dependent analysis for refused admissions in clinical wards, Annals of Operation Research, 178, 45-65.

[12] Bekker, R., Paulien, M. Koeleman (2011) Scheduling admissions and reducing variability in bed demand, Health Care Management Science, DOI 10.1007/s10729-011-9163-x.

[13] Bell. C., Allen. D (1969) Optimal planning of an emergency ambulance service. Socio and Economic Planning Sciences, 3, 95 .

[14] Biju M. K, Naeema, K., Faisal U (2011) Application of Queueing Theory in Human Resource Management in Health Care, ICOQM-10,June 28-30, 1019-1027.

[15] Blair, E.L., Lawrence, C.E (1981) A queueing network approach to health care planning with an application to burn care in New York state, Socio-economic Planning Sciences, 15, 207-216.

[16] Boyce, T., Hare, S. (1998) Effect of a computerised prescription log on prescription turnaround time in a hospital pharmacy. Pharmacy Management, 14(2), 17-20.

[17] Brahimi, M., Worthington, D.J.(1991) Queueing models for out-patient appointment systems? a case study, The Journal of the Operational Research Society, 42, 733-746.

[18] Broyles, J.R., Cochran, J.K. (2007) Improving patient safety by maximizing fast-track benefits in the ED-a queuing network approach. In: IIE industrial engineering research conference, Memphis, TN, 2007. 6pp. on CD-ROM.

[19] Broyles, J.R., Roche, Kevin, T. (2007) Capacitating Outpatient Clinics Using Queuing Theory, IIE Annual Conference. Proceedings, January 1, (2008) Industial Engineering Research Conference, 613-618.

[20] De Bruin, A.M., Koole, G.M., Visser, M.C (2005) Bottleneck analysis of emergency cardiac in-patient flow in a university setting: an application of queueing theory. Clinical and Investigative Medicine, 28, 316-317.

[21] De Bruin, A.M., Rossum, A.C., Visser, M.C., Koole, G.M (2007) Modelling the emergency cardiac in-patient flow: an application of queueing theory, Health Care Management Science, 10, 125-137.

[22] De Bruin, A.M., Bekker, R., Van Zanten, L., Koole, G.M. (2010) Dimensioning hospital wards using the Erlang loss model, Annals of Operation Research, 178, 23-43.

[23] Burwell, T.H., Mcknew, M.A. and Jarvis, J.P. (1992) An application of a spatially distributed queueing model to an ambulance system, Socio-Economic Planning Sciences, 26, 4, 289-300.

[24] Burwell, T.H., Jarvis, J.P., Mcknew, M.A (1993) Modelling co-located servers and dispatch ties in the hypercubemodel, Computers and Operations Research, 20(2), 113-119.

[25] Bretthauer, K.M., Sebastian Heese, H., Hubert Pun, Edwin Coe (2011) Blocking in Healthcare Operations: A New Heuristic and an Application, Production and Operations Management, 20(3), 375-391.

[26] Carri Chan., Galit Yom-Tov (2011) Intensive Care Unit Patient Flow with Readmissions: A State-Dependent Queueing Network,2011 MSOM Annual Conference Ann Arbor, Michigan,1-3, June 26-28.

[27] Chaussalet T.J., Xie H., Millard P (2006) A Closed Queueing Network Approach to the Analysis of Patient Flow in Health Care Systems, Methods Information Medicine, 5.

[28] Cochran, J.K., Bharti, A (2006) A multi-stage stochastic methodology for whole hospital bed planning under peak loading, International Journal of Industrial and Systems Engineering, 1(1), 8-36.

[29] Cochran, J.K., Bharti, A. (2006b) Stochastic bed balancing of an obstetrics hospital, Health Care Management Science, 9(1), 31-45.

[30] Cochran, J.K., Roche, K.T (2008) A queueing-based decision support methodology to estimate hospital inpatient bed demand, Journal of the Operational Research Society, 59, 1471-1482.

[31] Cochran, J.K., Roche, K.T (2009) A multi-class queueing network analysis methodology for improving hospital emergency department performance, Computers and Operations Research, 36, 1497-1512.

[32] Cochran, J.K., Broyles J.R (2010) Developing nonlinear queueing regressions to increase emergency department patient safety: Approximating reneging with balking, Computers and Industrial Engineering, 59, 378-386.

[33] Collings, T., Stoneman, C (1976) The $M/M/\infty$ Queue with varying arrival and departure rates, Operations Research, 24(4), 760-773.
[34] Cooper, J.K., Corcoran, T.M (1974) Estimating bed needs by means of queueing theory, The New England Journal of Medicine, (1974), 291, 404-405
[35] Creemers Stefan., Lambrecht Marc., Vandaele (2007) Queueing Models in Health care, Tijdschrift voor Economie en Management, Vol. LII, 3, 471-497.
[36] Creemers Stefan., Lambrecht Marc (2011) The Modelling of Interrupts and Unplanned Absences in Health Care Operations, Supply Chain Forum An International Journal Vol.12-N1.
[37] Eugene Day, T, Max Li, W., Armann Ingolfsson, Nathan Ravi (2010) The Use of Queueing and Simulative Analyses to Improve an Overwhelmed Pharmacy Call Center, Journal of Pharmacy Practice, 23(5) 492-495.
[38] van Dijk, N (2000) On hybrid combination of queueing and simulation. In Proceedings of the 2000 Winter simulation Conference, pages 147-150.
[39] van Dijk, N. M. and van der Sluis, E (2008) Practical optimisation by OR and simulation. Simulation Modelling Practice and Theory, 16(8), 1113-1122. EUROSIM 2007.
[40] Donehew, G.R., Hammerness, F.C (1978) How to measure time that is involved in filling Rxs.Pharm Times, 44, 54-59.
[41] DeLaurentis, P., Kopach, R., Rardin, R., Lawley, M., Muthuraman, K., Wan, H., Ozsen, L. and Intrevado, P (2006) Open access appointment scheduling an experience at a community clinic, in IIE Annual Conference and Exposition.
[42] E. El-Darzi A, C. Vasilakis B, T. Chaussalet A And P.H. Millard, A (1998) Simulation Modelling Approach To Evaluating Length Of Stay, Occupancy, Emptiness And Bed Blocking In A Hospital Geriatric Department, Health Care Management Science 1, 143-149.
[43] Fiems, D., Koole, G., Nain, P. (2007) Waiting times of scheduled patients in the presence of emergency requests. http://www.math.vu.nl/koole/articles/report05a/art.pdf, accessed August 6, (Working paper)
[44] Fletcher, A., Halsall, D., Huxham, S., Worthington, D. (2007) The DH accident and emergency department model: A national generic model used locally, Journal of Operational Research Society 58, 1554â€“1562.
[45] Michael Foster., Michael R Hosking., Serhan Ziya (2010) A Spoonful of Math Helps the Medicine Go Down: An Illustration of How Health care can Benefit from Mathematical Modelling and Analysis, BMC Medical Research Methodology, 10:60, http://www.biomedcentral.com/1471-2288/10/60.
[46] Goddard, J., Tavakoli, M (2008) Efficiency and welfare implications of managed public sector hospital waiting lists, European Journal of Operational Research, 184, 2, 778-792.

[47] Gorunescu, F., McClean, S.I., Millard, P.H (2002) Using a Queueing Model to Help Plan Bed Allocation in a Department of Geriatric Medicine, Health Care Management Science, 5, 307-312.

[48] Gorunescu, F., McClean, S.I., Millard, P.H (2002) A queueing model for bed-occupancy management and planning of hospitals, Journal of the Operational Research Society, 53, 19-24.

[49] Gonzalezb Paula., Herrero Carmen (2004) Optimal sharing of surgical costs in the presence of queues, Mathematical Methods of Operation Research, 59, 435-446.

[50] Green, L. V.(2002) How many hospital beds? Inquiry, 39(4) 400-412.

[51] Green, L.V. (2006) Queueing analysis in health care, in Patient Flow: Reducing Delay in Health care Delivery, Hall, R.W., ed., Springer, New York, 281-308.

[52] Green, L.V., Savin, S (2008) Reducing delays for medical appointments: A queueing approach. Operations Research, 56(6), 1526-1538.

[53] Green, L.V., Armand G. (2010) Using Queueing theory to alleviate emergency department overgrowing, Wiley Encyclopedia of Operations Research and Management Science, edited by James J. Cochran, 1-9.

[54] Green, L. V., Nguyen, V (2001) Strategies for cutting hospital beds: the impact on patient service. Health Services Research, 36, 421-442.

[55] Green, L.V., Soares, J., Giglio J.F., Green R.A. (2006) Using queueing theory to increase the effectiveness of emergency department provider staffing, Academic Emergency Medicine, 13, 61-68.

[56] Green L.V., Kolesar, P.J., Whitt, W. (2007) Coping with time-varying demand when setting staffing requirements for a service system, Product Operation Management, 16: 13-39.

[57] Griffths.J.D, Price-Llyod N (2006) A Queueing Model of activities in an intensive care unit, IMA Journal of Management Mathematics, 17, 277-288.

[58] Gross and Harris (1998) Fundamentals of queueing theory. Wiley-India.

[59] Gupta,I., Zoreda, J., Kramer, N. (1971) Hospital manpower planning by use of queueing theory. Health Services Research, 6, 76-82.

[60] Hall, R.W (2006) Patient Flow: Reducing Delay in Health care Delivery, first edn. Springer, New York.

[61] Hall, R.W (2006) patient flow: The new queueing theory for healthcare OR/MS Today, 23(3), 36-40.

[62] Haussmann, R.K.D (1970) Waiting time as an index quality of nursing care. Health Services Research, 5, 92-105.

[63] Jeffrey W. Herrmann (2008) Queueing network approximations for Mass Dispensing and Vaccination Clinics, ISR Technical Report, 2.

[64] Ho, C.J, Lau, H. S (1992) Minimizing total cost in scheduling outpatient appointments, Management Science, 38(12), 1750-1764.

[65] Huang, X.M (1994) Patient attitude towards waiting in an outpatient clinic and its applications, Health Services Management Research, 7, 1, 2-8.
[66] Iannoni, A.P., Morabito, R (2007) A multiple dispatch and partial backup hypercube Queueing model to analyse emergency medical systems on highways, Transportation Research, Part E, 43, 755-771.
[67] Igor Georgievskiy, Zhanna Georgievskaya, William Pinney, Donald Mcwilliams (2008), Using Queueing Analysis And Computer Simulation Modeling To Reduce Waiting Time In The Hospital Admitting Department
[68] Igor Georgievskiy., Zhanna Georgievskaya., Wm Pinney (2009) An application of Queueing analysis to the waiting room problem in the hospital admission process, 181-189, Conferene proceedings.
[69] Jackson, R.R.P., Welch, J.D., Fry, J. (1964) Appointment systems in hospitals and general practice. Operation Research Quartly, 15, 219-237.
[70] Jiang, L., Giachetti, R. (2008) A queueing network model to analyse the impact of parallelization of care on patient cycle time. Health Care Management Science, 11(3), 248-261.
[71] John Kolade Obamiro (2010), Queueing Theory and Patient Satisfaction: An Overview of Terminology and Application in Ante-Natal Care Unit, Economic Sciences Series, Vol. LXII No. 1
[72] Jones, R (2011) Hospital bed occupancy demystified. British Journal of Health care Management, 17(6), 242-248.
[73] Joustra, P., van der Sluis, E., van Dijk, N.M (2010) To pool or not to pool in hospitals: a theoretical and practical comparison for a radiotherapy outpatient department, Annals
[74] Jun, J.B., Jacobson, S.H., Swisher, J.R (1999) Application of discrete-event simulation in health care clinics, A survey. Journal of the Operational Research Society, 50(2), 109-123.
[75] Kao, E.P.C., Tung, G.G (1981) Bed allocation in a public health care delivery system, Management Science, 27, 507-520.
[76] Kapadia A. S, Chiang. Y. K, Kazmi. M. F (1985) Finite capacity priority queues with potential health applications, Computer and Operation Research, (4), 411-420.
[77] Karnon, J., Mark Mackay., Mills, T.M (2009) Mathematical modelling in health care, 18th World IMACS / MODSIM Congress, Cairns, Australia, 13-17.
[78] Keller, T.F., Laughhunn D.J (1973) An application of queueing theory to a congestion problem in an outpatient clinic, Decision Sciences, 4, 379-394.
[79] Khan, M.R., Callahan, B.B (1993) Planning laboratory staffing with a queueing model, European Iournal of Operational Research, 67, 321-331.
[80] Kim, S.C., Horowitz, I., Young, K.K., Buckley, T.A (1999) Analysis of capacity management of the intensive care unit in a hospital, European Journal of Operational Research, 115(1), 36-46.

[81] Koizumi, N., Kuno, E., Smith, T.E (2005) Modelling patient flows using a queueing network with blocking, Health Care Management Science, 8, 49-60.
[82] Ladany, S.P., Turban, E (1978), A Simulation Of Emergency Rooms, Computers and Operation Research, 4, 89-100.
[83] Larson, R.C (1974) A hypercube queueing model for facility location and redistricting in urban emergency services, Computers and Operations Research, 1, 67-95.
[84] Lin, A.C., Jang, R., Sedani, D (1996) Re-engineering a pharmacy work system and layoutto facilitate patient counseling, American Journal of Health System Pharmacy, 53,1558-1564.
[85] Lin, A.C, Jang, R., Lobas, N (1999) Identification of factors leading to excessive waiting times in an ambulatory pharmacy, Hospital Pharmacy, 34, 707-712.
[86] Liyanage, L., Gale M (1995) Quality Improvement for the Campbell Town Hospital Emergency Service. In: 1995 IEEE International Conference on Systems, Man, and Cybernetics. Institute of Electrical and Electronics Engineers, Vancouver, British Columbia, Canada, 22-25 , October 13, 1997-2002.
[87] Mandelbaum Avishai el (2010) On fair routing from emergency departments to hospital wards: QED queues with heterogeneous servers, Working paper.
[88] Mayhew, L., Smith.D (2008) Using queueing theory to analyse the Governmentos 4-h completion time target in Accident and Emergency departments, Health Care Management Science, 11, 11-21.
[89] McClain, J.O (1976) Bed planning using queueing theory models of hospital occupancy: a sensitivity analysis, Inquiry, 13, 167-176.
[90] McManus, M.L., Long, M.C., Cooper, A., Litvak, E (2004) Queueing theory accurately models the need for critical care resources. Anesthesiology, 100, 1271-1276.
[91] McQuarrie D.G (1983) Hospital utilisation levels. The application of queueing theory to a controversial medical economic problem, Minnesota Medicine, 66, 679-86.
[92] Mendonca, F.C., Morabito, R (2001) Analysing emergency medical service ambulance deployment on a Brazilian highway using the hypercube model, Journal of the Operation Research Society, 52, 261 -270.
[93] Milliken, R.A., Rosenberg, L., Milliken, G.M (1972) A queueing model for the prediction of delivery room utilisation, American Journal of Obstetrics and Gynecology, 114, 691-699.
[94] Moore, B.J (1977) Use of queueing theory for problem solution in Dallas, Tex., Bureau of Vital Statistics, Public Health Reports, 92, 171-175.
[95] Mokaddis, G.S., Ismail, I.A., Metwally, S.A., Metry, K.M (2011) Response Times for Health care System, Journal of Applied Mathematics and Bioinformatics, 1(2), 131-146.
[96] Moss, G (1987) Hospital pharmacy staffing levels and outpatient waiting times, The Pharmacy Journal, 239(Jul 18), 69-70.

[97] Marek Laskowski, Robert D. McLeod, Marcia R. Friesen, Blake W. Podaima, Attahiru S. Alfa (2009) Models of Emergency Departments for Reducing Patient Waiting Times, July PLoS ONE, 4(7), 1-12.

[98] Natalia Yankovic, Green.L.V (2011) Identifying Good Nursing Levels: A Queueing Approach, Operation Research, 59(4), 942-955.

[99] Ndukwe H. C., Omale S., Opanuga O (2011) Reducing queues in a Nigerian hospital pharmacy, African Journal of Pharmacy and Pharmacology, 5(8), 1020-1026.

[100] Nosek, Jr., R.A., Wilson, J.P. (2001) Queueing theory and customer satisfaction: a review of terminology, trends, and applications to pharmacy practice. Hospital Pharmacy, 36, 275-279.

[101]Carolina Osorio, Michel Bierlaire (2009) An analytic finite capacity queueing network model capturing the propagation of congestion and blocking ,European Journal Of Operational Research, 196(3), 996-1007.

[102]Pajouh, Foad Mahdavi, Kamath, Manjunath (2010) Applications of Queueing Models in Hospitals, MWAIS 2010 Proceedings. 23, http://aisel.aisnet.org/mwais2010/23.

[103] Palvannan, R.K, Teow, K.L (2010), Queueing for Health care, Doi 10.1007/S10916-010-9499-7.

[104] Panayiotopoulos, J.C., Vassilacopoulos, G (1984) Simulating hospital emergency departments queueing systems, European Journal of Operational Research, 18, 250-258.

[105] Preater, J (2001) A bibliography of queues in health and medicine, Keele Mathematics Research Report.

[106] Preater, J (2002) Queues in health. Health Care Management Science, 5, 283.

[107] Reetu Mehandiratta (2011) Applications of Queueing theory in health care, International Journal of Computing and Business Research, Volume 2 Issue 2, ISSN: 2229-6166.

[108] Mateo Restrepo, Shane G. Henderson, Huseyin Topaloglu (2009) Erlang loss models for the static deployment of ambulances: Health Care Management Science, 12, 67-79.

[109] Ridge, J.C., Jones, S.K., Nielsen, M.S., Shahani, A.K (1998) Capacity planning for intensive care units, European Journal of Operational Research, 105, 346-355.

[110] Roche, K. T., Cochran, J.K (2007) Improving patient safety by maximizing fast track benefits in the ED ? A queueing network approach. In IIE Industrial Engineering Research Conference (6p), Nashville, TN [CD-ROM].

[111] Rosenquist, C.J (1987) Queueing analysis: a useful planning and management technique for radiology, Journal of Medical Systems, 11, 413-419.

[112] Samuel Fomundam, Jeffrey Herrmann (2007) A Survey of Queueing Theory Applications in Health care, ISR Technical Report, 24.

[113] Seshaiah C.V., Thiagaraj, H.B (2011) A Queueing Network Congestion Model in Hospitals, European Journal of Scientific Research, 63(3), 419-427.

[114] Scott, D.W, Lynette E. Factor, G. Anthony Gorry, (1978) Predicting the response time of an urban ambulance system, Health service research, 404-417.

[115] Shimshak, D.G., Gropp Damico, D., Burden, H.D (1981) A priority queueing model of a hospital pharmacy unit, European Journal of Operational Research, 7, 350-354.

[116] Shmueli, A., Sprung, C.L., Kaplan E.H. (2003) Optimizing admissions to an intensive care unit, Health Care Management Science, 6(3), 131-136.

[117] Siddharthan., Kris., Jones Walter J (1996) A priority queueing model to reduce waiting times in emergency care, International Journal of Health Care Quality Assurance, 10-16, 9/5.

[118] Singer, M., Donoso, P (2008) Assessing an ambulance service with queueing theory, Computers and Operation Research, 35, 2549-2560

[119] Solberg, L., Asplin, B., Weinick R., Magid, D (2003) Emergency department crowding: consensus development of potential measures, Annals of Emergency Medicine, 42, 6, 824-834.

[120] Taylor, T.H., Jennings, A.M.C., Nightingale, D.A., Barber, B., Leivers, D., Styles, M., Magner, J. (1969) A study of anaesthetic emergency work, Paper 1: The method of study and introduction of queueing theory, British Journal of Anaesthesia, 41, 70-75.

[121] Taylor. I.D.S., Templeton J.G.C (1980) Waiting time in a multi-server cutoff-priority queue, and its application to an urban ambulance service, Operations Research, 28(5), 168-1188.

[122] Tucker, J.B., Barone, J.E., Cecere, J., Blabey, R.G., Rha, C (1999) Using queueing theory to determine operating room staffing needs, Journal of Trauma, 46, 71-79.

[123] Vanberkel, P.T., Boucherie, R.J., Hans, E.W., Hurink, J.L., Litvak, N (2010) A survey of health care models that encompass multiple departments, International Journal of Health Management and Information, 1(1), 37-69.

[124] Vasanawala, S.S., Desser, T.S (2005) Accommodation of requests for emergency US and CT: Applications of queueing theory to scheduling of urgent studies, Radiology, 235, 244-249.

[125] Vemuri, S (1984) Simulated analysis of patient waiting time in an outpatient pharmacy, Am erican Journal Hospital Pharmacy, 41, 1127-1130. 27Q-Matic System company brochure.Ashville, NC: The Q-Matic Corporation.

[126] De Vericourt, F., Jennings, O.B (2011) Nurse Staffing in Medical Units:A Queueing Perspective, Operation Research, 59(6), 1320-1331.

[127] Vikas Singh, (2006) Use of Queueing Models in Health Care, University of Arkansas for Medical Sciences, University of Arkansas for Medical Sciences, 58.

[128] Viswanadham, N., Narahari, Y (2001) Queueing network modelling and lead-time compression of pharmaceutical drug development, International Journal of Production Research., 39, 2, 395-412.

[129] Juyun Wang., Hua Yu.,Jianyu Luo., Jie Sui (2008) Medical Treatment Capability Analysis using Queueing Theory in a Biochemical Terrorist Attack, The 7th International Symposium on Operations Research and Its Applications (ISORA08) ijiang, China, 415-424, October 31-Novemver 3.
[130] Weiss, E.N, McClain, J.O (1987) Administrative days in acute care facilities, A queueing-analytic approach, Operations Research 35(1), 35â€“44.
[131] Worthington, D.J(1987) Queueing Models for Hospital Waiting Lists, The Journal of the Operation Research Society, 38, 413-422.
[132] Worthington, D.J (1991) Hospital waiting list management models, Journal of the Operational Research Society, 42(10), 833-843.
[133] Xiaodong Li, Patrick Beullens, Dylan Jones, Mehrdad Tamiz (2009) Optimal Bed Allocation in Hospitals, Lecture Notes in Economics and Mathematical Systems, 618, 253-265.
[134] Young J.P (1962) The basic models, in A Queueing theory approach to the control of hospital inpatient census, John Hopkins University, Baltimore, 74-97.
[135] Young J.P (1962b) Estimating bed requirements, in A Queueing theory approach to the control of hospital inpatient census, John Hopkins University, Baltimore, 98-108.
[136] Adrian H. Zai, Md, Phd, Mph, Kit M. Farr, Md, Richard W. Grant, Md, Mph, Elizabeth Mort, Md, Mph, Timothy G. Ferris, Md, Mph, Henry C. Chueh, Md Ms (2009) Queueing Theory to Guide the Implementation of a Heart Failure Inpatient Registry Program, J Am Med Inform Assoc.16:516-523. DOI 10.1197/jamia.M2977
[137] Zeltyn, S., Carmeli B., Greenshpan O., Mesika Y., Wasserkrug S., Vortman P., Marmor Y.N., A. Mandelbaum, A. Shtub, T. Lauterman D. Schwartz, K. Moskovitch, S. Tzafrir, F. Basis (2009) Simulation-based models of emergency departments: Operational, tactical and strategic staffing. Submitted to ACM Transactions on Modelling and Computer Simulation (TOMACS).
[138] Zenios, S.A (1999) Modelling the transplant waiting list: A queueing model with reneging, Queueing Systems, 31, 239-251.
[139] Zenios, S.A , Chertow, G.M and Wein, L.W (2000) Dynamic Allocation Of Kidneys To Candidates On The Transplant waiting List, Operations Research, 48(4), 549-569
[140] Zhao, L., Lie, B (2008) Modeling and Simulation of Patient Flow In Hospitals for Resource Utilization, Conference Proceeding.
[141] Zonderland, M.E., Boer, F., Boucherie, R.J., de Roode, A., van Kleef, J. (2009) Redesign of a university hospital preanesthesia evaluation clinic using a queueing theory approach, Anesthesia and Analgesia, 109(5), 1612-1621.