KU LEUVEN

FACULTY OF ECONOMICS AND BUSINESS

Improving patient flow in emergency departments with OR techniques: a literature overview

Carmen R, Van Nieuwenhuyse I.

KBI_1425

Improving patient flow in emergency departments with OR techniques: A literature overview

Raïsa Carmen* and Inneke Van Nieuwenhuyse
Research Center for Operations Management
Department of Decision Sciences and Information Management
KU Leuven, Belgium
raisa.carmen@kuleuven.be
inneke.vannieuwenhuyse@kuleuven.be

November 10, 2014

1 Introduction

The emergency department (ED) is an interesting field for operations research (OR) and operations management (OM) researchers. Having time-varying arrivals and heterogeneous patients that need to be treated in consecutive processing steps by several doctors, nurses and other employees, it is a complex environment to control. The difficulty to control EDs often results in *(over)crowded* EDs. In general, an ED is said to be experiencing periods of crowding when the demand for ED services exceeds the available ED resources (Higginson, 2012). ED crowding is a worldwide phenomenon with regional influences (Jayaprakash et al., 2009) and may cause a myriad of operational problems like patients being treated in hallways, excessive length of stay (LOS) and wait times, patients leaving without treatment or even medical errors and increased mortality. Descriptive studies that define crowding (Hwang & Concato, 2004), examine the causes and effects (Richardson & Mountain, 2009; Hoot & Aronsky, 2008; Asplin et al., 2003), and propose models to measure crowding (Hwang et al., 2011; Higginson, 2012) can be found in the medical literature. Applying OR and OM techniques to improve the performance of EDs is needed now more than ever since health budgets are tight, demand for health care services is rising and higher performance standards are being demanded simultaneously.

*Corresponding author.

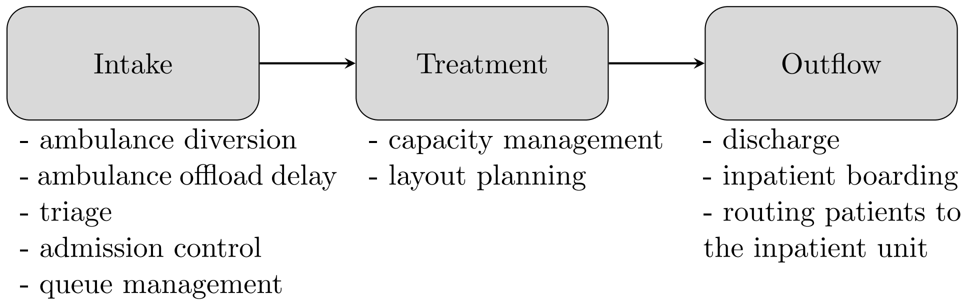

Figure 1: The ED structure

The scope of the articles that have been selected for discussion and the search method are delineated in Section 1.1 and Section 1.2 respectively. Section 2 will provide an overview of the selected articles that use OR and OM techniques to improve ED processes. Considering the existing literature, Section 3 will outline my future research perspectives.

1.1 Scope

Inspired by Asplin et al. (2003), Saghafian et al. (2014a), and Welch (2012) that all split up the ED in three parts, the articles will be structured according to the framework of Figure 1; the intake-treatment-outflow model. The intake and outflow of the ED have received less interest in the OR/OM literature. In practice, however, the intake and outflow have been identified as 'problem areas', in need of well-founded improvements (Crawford et al., 2013; Hall, 2006; Richardson & Mountain, 2009; Welch & Savitz, 2012; Wiler et al., 2010). Since good literature reviews already exist on capacity management (Defraeye & Nieuwenhuyse, 2013; Saghafian et al., 2014a), we will thus focus on the first and last step where the ED interacts with its environment.

We aim our attention at state-of-the-art articles, published mainly after 2005. However, if the OR/OM literature is too sparse on a certain topic, we refer to less recent articles or medical journal articles describing the problem at hand from a practice perspective.

1.2 Search method

The first step in finding relevant articles within the scope outlined above was a broad search in the well known Web of Science database using the general keywords of Table 1. Web of Science was chosen since this database

General keywords	"emergenc*", "acute", "accident", "health", "patient flow", "hospital"
Additional intake keywords	"diversion", "bypass", "triage", "waiting", "offload", "ramping"
Additional outflow keywords	"discharge", "board*", "access block", "block*"

Table 1: Search terms used

allows searching in all ISI-listed journals. The second step was a more detailed search in the journal archives of the journals listed in Table 2. Here, some additional keywords, listed in Table 1, were used to find articles that specifically focused on the intake or outflow of the ED. Articles that were published before 2005 and did not apply OR techniques or did not focus on the ED were immediately eliminated. Next, the selected articles were reviewed in detail and classified. A backward and forward citation search on the key articles was performed to identify important articles published before 2005, new working papers or 'articles in advance'.
In addition to the OR/OM articles, we will also refer to articles originating from medical journals, albeit in less detail. The most often cited medical journals are *Academic Emergency Medicine* (17 articles), *Annals of Emergency Medicine* (14 articles), *Emergency Medicine Journal* (7 articles) and *The Journal of Emergency Medicine* (7 articles).

Frequently used OR/OM techniques in health care applications are queueing theory (C & Iyer, 2013; Fomundam & Herrmann, 2007; Green, 2006) and simulation (Günal & Pidd, 2010; Katsaliaki & Mustafee, 2011; Paul et al., 2010); these methods will often be mentioned in the literature review. For a detailed description of other OR/OM techniques used in EDs, we refer the interested reader to Bhattacharjee & Ray (2014), Lim et al. (2012), Marshall et al. (2005), and Saghafian et al. (2014a).

Journal	Nr. of articles
Computers & Operations Research	1
Decision Support Systems	1
European Journal of Operational Research	9
Expert Systems with Applications	4
Health Care Management Science	9
Health Systems	0
IIE Transactions	0
International Journal of Operations & Production Management	1
International Journal of Production Economics	3
International Journal of Production Research	0
Journal of Operations Management	0
Journal of the Operational Research Society	4
Management Science	11
Manufacturing & Service Operations Management	9
Omega - International Journal of Management Science	2
OR Spectrum	0
Operations Research	9
Operations Research for Health Care	4
Production & Operations Management	4
Proceedings of the Winter Simulation Conference	13
Total	84

Table 2: Journals used in detailed search, with the number of articles selected

2 Literature review

This section gives an overview and brief discussion of all relevant articles. In total, we refer to almost 100 OR articles of which 72 will be highlighted and classified in tables. Section 2.1 discusses the articles that focus on the inflow of patients in the ED. Section 2.2 discusses how the outflow of patients from the ED can be improved.

2.1 Intake

Too many patients leave the ED before they have been seen by a physician. This is referred to as *leave without being seen (LWBS)* or *leave without treatment (LWOT)* (Kennedy et al., 2008). To avoid this, a smooth intake of patients in the ED is necessary. Welch & Savitz (2012) list several innovations that are used in practice to improve the ED intake and categorize them into three categories; physical plant changes, technological changes, and process/flow changes. The OR literature focuses strongly on process/flow changes.

Wiler et al. (2010) describe the ED 'front-end' as the patient care processes that occur from the time of a patient's initial arrival to the ED to the time an ED health care provider formally assumes responsibility for the comprehensive evaluation and treatment of the patient. Before we discuss articles that focus on triage (Section 2.1.2) and queue management (Section 2.1.3), we first consider OR articles that consider admission control and the interface with the emergency medical service providers in Section 2.1.1.

2.1.1 Admission control

Unlike production environments where demand and supply can be somewhat manipulated, smoothing the demand for emergency services is almost impossible in practice. Predicting the demand is hard and denying emergency care is not even allowed in most countries. In medical literature, there have been quite some studies about the inappropriate use of the ED by patients with minor injuries (Carret et al., 2009). However, there is insufficient evidence that these patients significantly affect crowding (Richardson & Mountain, 2009) and, from an OR/OM perspective, little can be done to avert these low-acuity patients. Nevertheless, methods exist to safely guide some patients away from the ED; Xu & Chan (2014) propose to internally divert patients to a different department or medical resource within the same hospital and Helm et al. (2011) advise to ask low-acuity patients to wait at home until they are called in when the ED is less busy. The most often used method to control ED admissions, however, is ambulance diversion.

Ambulance Diversion (AD) or *ambulance bypass* was first introduced by Lagoe & Jastremski (1990) as a new strategy to alleviate crowding in hospital EDs. The policy consists of rerouting incoming ambulances to neighboring hospitals in periods of crowding. Over the years, the conjecture that AD may have undesirable consequences gave rise to a stream of articles investigating possible causes and effects of ambulance diversion (AD). Pham et al. (2006) give an overview of these articles and conclude that AD seems to have a small adverse impact on transport and treatment times. However, there is no decisive evidence that AD affects mortality or the financial health of the hospital. Since AD usually coincides with crowding, disentangling AD consequences from the crowding effects (and thus also proving the existence

Topic	Research goal			Article	Number of hospitals	Diversion policy based on	Methodology			Extensions
	Forecasting	System evaluation	Policy optimization				Simulation	Queueing	Other	
AD	X			Au et al. (2009)	1	ED status		X		IU
AD	X			Chockalingam et al. (2010)	1	ED + IU status	X		PN	IU
AD		X		Ramirez et al. (2009)	1	ED status	X			IU (+DA)
AD		X		Ramirez-Nafarrate et al. (2010)	1	ED + IU status	X			IU (+DA)
AD		X		Allon et al. (2013)	1	ED status	X	X		IU (+DA)
AD			X	Ramirez-Nafarrate et al. (2014)	2	ED status	X	X		IU (+DA)
AD			X	Ramirez-Nafarrate et al. (2011)	3	ED + IU status	X		GA	IU (+DA)
AD		X	X	Xu & Chan (2014)	1	ED status	X	X		
AD		X	X	Deo & Gurvich (2011)	2	ED status	X	X	GT	
AD		X	X	Hagtvedt et al. (2009)	2-N	ED status	X	X	GT	
AOD		X		Almehdawe et al. (2013)	N	/	X	X		EMS provider
AP			X	Helm et al. (2011)	N	/	X	X		

Table 3: Overview of admission control articles

of a relationship between AD and these negative consequences) is difficult.

OM and OR research on the AD phenomenon has been increasing in recent years. Delgado et al. (2013) review ten articles that incorporate, investigate and suggest strategies to avoid the AD problem using simulation. We used a slightly different classification and selection of AD articles than Delgado et al. (2013) and uncovered three possible research goals in the context of AD; forecasting the probability of AD in real time, system evaluation, and policy optimization. Table 3 gives an overview of the relevant articles that will be discussed here.

The diversion policy, which specifies when to divert or accept an incoming ambulance, is based on the ED status and/or the inpatient unit (IU) status. If the diversion policy is based on the ED status, this means that there is a threshold on either the number occupied ED beds, the total number of patients (waiting) in the ED or the number of patients boarding (we refer to Section 2.2 for more details on boarding). Thresholds on the IU status are typically based on the number of occupied IU beds.

The goal of Au et al. (2009) and Chockalingam et al. (2010) is to find a way to dynamically predict the probability that AD will be necessary in the near future. These predictions could then be used in operational decision making to bring in additional resources (Au et al., 2009; Chockalingam et al., 2010) or free inpatient beds (Au et al., 2009), thus avoiding or limiting ambulance diversions. Chockalingam et al. (2010) use Petri-nets (PN) to model patient and resource flow in a hospital and derive the underlying stochastic control problem to determine exactly which resources need to be added or removed to bring the *distance to divert* (a measure of the proximity of a hospital to a divert state) back to a safe level when needed. Similarly, Epstein & Tian (2006) develop a work score based on the number of patients in the waiting room, their triage level, and the number of boarders. Their marginal multivariate logistic regression model helps to predict AD and to make more objective decisions on when to go on diversion.

Allon et al. (2013), Ramirez et al. (2009), and Ramirez-Nafarrate et al. (2010) perform what-if analyses on an ED system with AD; they analyze and evaluate the phenomenon and its sensitivity to changing ED characteristics and environment. Ramirez et al. (2009) vary the diversion threshold on the number of patients waiting for treatment in a simulation model and evaluate the effect of these variations on the total number of patients treated, the waiting time of patients, the utilization of the ED beds, LWOT, and the percentage of time on diversion. They conclude that there is a trade-off between diversion and the other ED performance measures such as waiting times and LWOT; the more diversion, the lower the waiting times and

LWOT rates. Ramirez-Nafarrate et al. (2010) consider policies based on the total number of patients waiting for a bed, the number of patients boarding in the ED, or the number of beds available in the IU. In addition to the *'regular'* policy with just one threshold, they also consider policies with two threshold parameters; one to start diverting and another one to stop diverting ambulances. Reevaluation of the AD status can either happen periodically or continuously. All policies are compared and evaluated based on the trade-off they achieve between timely service and accessibility. Using diffusion and fluid approximations of a multidimensional Markov process that are validated with respect to empirical evidence, Allon et al. (2013) derive that the time that the ED spends on diversion is lower for larger EDs with more spare capacity in the IU and/or less neighboring hospitals.

The policies explored in Ramirez-Nafarrate et al. (2011) are similar to those in Ramirez-Nafarrate et al. (2010) but Ramirez-Nafarrate et al. (2011) search for the *optimal* policy that minimizes the non-value added time (boarding, waiting, and transport time) using a genetic algorithm (GA). Additionally, they test if sending patients to the least crowded hospital could give better results compared to sending them to the nearest hospital and if the relative size of the regional hospitals matters. Ramirez-Nafarrate et al. (2014) aim to minimize the time that patients wait beyond their *recommended safety time threshold (RSTT)*. The ED is split up in an urgent and a less-urgent area. Using a Markov Decision Process, they search for optimal AD control policies that not only take into account the number of patients but also the severity mix of the patients in the ED. They assess the sensitivity of the thresholds to the patient traffic and the severity mix and examine whether information on the severity level of an incoming ambulance patient, or on the time to start treatment in the neighboring hospitals can improve AD decision making.

Xu & Chan (2014) use forecasts of the expected demand in their proactive diversion policy. By not just waiting until the ED is full but proactively diverting when the ED is *expected* to be full in the near future, smaller waiting times can be achieved with the same level of diversion. Their policy has two thresholds on the number of patients waiting for treatment; a lower threshold under which no diversion is allowed and an upper threshold above which all patients are diverted. If the number of patients waiting lies between the two thresholds, patients are diverted with a probability that increases as the number of patients waiting increases.

Deo & Gurvich (2011) and Hagtvedt et al. (2009) take into account that patients who are diverted from one hospital will need to be admitted to another hospital. They reason that the *socially optimal* policy takes into account all patients from all hospitals and use game theory (GT) to demonstrate that hospitals will very likely make selfish, sub-optimal diversion decisions if there is no central decision-making body or *social planner*. While Hagtvedt et al. (2009) focus on minimizing the total time spent on diversion as much as possible, Deo & Gurvich (2011) acknowledge the potential of AD as a tool to alleviate crowding and aim to minimize the average waiting time across the entire network of hospitals. Deo & Gurvich (2011) pose that decentralized decision making might drive the hospitals to a *defensive equilibrium* where all hospitals refuse diversions from neighboring hospitals in an effort to minimize their own waiting time, thus effectively annihilating all possible pooling benefits.

Ambulance offload delay (AOD) or *ambulance ramping* occurs when an arriving ambulance cannot transfer patient care to the ED staff immediately because of insufficient ED resources (beds or staff). Cooney et al. (2013) concluded from an observational study that AOD, like AD, is associated with ED crowding. When patients are stranded on ambulance stretchers, they occupy precious ambulance resources which has obvious negative consequences for patient health, safety and hospital costs (Cooney et al., 2011, 2013). Articles on the topic of AOD mainly originate from medical journals where, for instance, the effect of adding dedicated ambulance *offload nurses* is quantified (Ovens, 2011) or the information transfer during patient handoff is studied in more detail (Jensen et al., 2013). Despite the fact that AOD has received quite some attention in medical journals, we found just one article that approached this problem from an OM perspective; Almehdawe et al. (2013) use a Markov chain to model the interaction between the EMS provider and the ED. The EMS provider owns several ambulances and is responsible for transporting patients to multiple hospitals within a certain region. Both walk-in and ambulance patients arrive to the EDs and ambulance patients always have preemptive priority over walk-in patients. The article develops algorithms that allow decision makers to easily compute the impact of changes in the number of beds in each of the hospitals on several system performance measures, such as offload delay for ambulance patients. The queueing network model is validated by simulation and appears quite robust to changes in the assumed transit times or ED service time distributions.

Asamoah et al. (2008) and Cooney et al. (2011) (both appearing in medical journals) investigate the link between AD and AOD. The former derived from a retrospective study that decreasing the time on diversion will likely increase crowding in the ED, thus raising the 'drop-off' time (i.e. offload delay). More specifically, by forcing a limit on the number of hours that a group of hospitals are allowed to be on diversion, they observed a decrease of 82% in the number of hours on diversion but an increase of 32% in the drop-off time. The latter suggests that, in extremely crowded EDs, AOD may be so large that it completely outweighs the slight increase in transportation times that AD may cause. Reducing crowding and consequently AOD by allowing AD may be a very attractive option in that case.

Helm et al. (2011) identify a third category of patients next to inpatients and emergency patients; the *expedited patients*. The acuity of their medical condition is less than most ED patients who are admitted, and their admission to the hospital can be delayed for a couple of days without compromising their health. Since the waiting list for scheduled elective admission is often too long for these patients, they usually enter through the ED. Helm et al. (2011), introduce an expedited call-in queue to cut down the load these patients put on the ED. Thus, hospital occupancy can by smoothed by calling in expedited patients whenever the hospital occupancy is low and canceling elective admissions when the occupancy is too high. The ability to smooth the hospital occupancy in this way will enhance the hospital performance. Using a Markov decision process (MDP), they seek the optimal admission policy (AP) that balances the opportunity cost of unutilized resources with the penalties associated with heavy congestion. The admission policy specifies when expedited patients can be called in or when scheduled patients need to be canceled, depending on the occupancy level and the number of patients in the call-in queue.

Nearly all authors use simulation in their research to evaluate the performance of a certain diversion policy or to validate their approximations (Table 3); the only exception is Au et al. (2009). Queueing also appears to be an accepted research method in these articles. Many of these articles in fact employ Markov processes (Allon et al., 2013; Almehdawe et al., 2013; Au et al., 2009; Deo & Gurvich, 2011; Hagtvedt et al., 2009; Xu & Chan, 2014) or even a Markov Decision process (Helm et al., 2011; Ramirez-Nafarrate et al., 2014). Other techniques that are being used are Petri-nets (PN), Game theory (GT) and genetic algorithms (GA) as described earlier in the text.

As can be seen in the last column of Table 3, seven out of the ten AD articles incorporate the inpatient unit (IU) and most of these articles also explicitly model direct admissions (DA) to this IU. The likely reasoning behind this is that AD is not just and ED problem, it is a hospital-wide problem (Millard, 2011). *Inpatient boarding* or *inpatient access block* causes patients to be *'blocked'* in the ED when there are not enough inpatient beds to accommodate all patients that need to be admitted. Inpatient boarding greatly influences ED crowding and will be discussed in further detail in Section 2.2.

2.1.2 Triage

Triage is typically the first step all patients must undergo, either in the ambulance on their way to the hospital or when they arrive in the ED. This process consists of sorting and prioritizing patients such that all patients get a clinically justified level of care and scarce resources are used efficiently (FitzGerald et al., 2010).

Table 4 gives an overview of the OR/OM articles that focus on triage or prioritization. Four research goals can be distinguished in this context: (1) the development of a new triage method, (2) the description of methods to uncover the hidden rules of triage, (3) the investigation of operational policies that could improve the flow through the triage system (and ultimately through the whole ED), and (4) the analysis of the trade-offs that should be made to establish if prioritization is beneficial or not.

Over the years, formalized triage scales that aid in making the triage process more repeatable and less subjective have been developed (FitzGerald et al., 2010). The most relevant triage scales in practice are Australian Triage scale (ATS), Canadian Triage and Acuity Scale (CTAS), Manchester Triage System (MTS), and Emergency Severity Index (ESI). Christ et al. (2010) give an overview of medical literature that contrast these triage scales with respect to validity and reliability. Apart from the four commonly used triage systems mentioned above, some slightly different systems have been introduced in OR journals. The triage system introduced in Saghafian et al. (2014b), called *complexity-augmented triage*, is based on both the urgency and the complexity of the patient's condition. The complexity is a function of the number of tests and treatment steps that the patient will require. As such, it represents the resource requirements of the patient. Although Vance & Sprivulis (2005) demonstrate that triage nurses are capable of assessing the patient's complexity in a reliable and valid way, Saghafian et al. (2014b) also test the prioritization system against misclassification errors. They conclude that complexity-augmented triage is relatively robust to misclassification error rates.

An interesting recent development in queueing theory that may be very useful for modeling the typical priorities in EDs more realistically is the accumulating priority queue of Sharif et al. (2014) and Stanford et al. (2014). In this queue, a patient's priority is a function of both the patient type and the time the patient has been waiting. While waiting, patients 'accumulate'

Research goal						Methodology		
(1)	(2)	(3)	(4)	Article	Focus	Simulation	Queueing	Other
X				Saghafian et al. (2014b)	complexity-augmented triage	X	X	
X				Sharif et al. (2014) and Stanford et al. (2014)	Accumulating priority queue	X	X	
X				Argon & Ziya (2009)	highest-signal-first policy		X	
X	X			Ashour & Kremer (2013)	FAHP-MAUT	X		
	X			Chonde et al. (2013)	predicting ESI scores			regression, artificial neural -and naive Bayesian networks
	X			Lin et al. (2010)	predicting a 4-level triage scale			cluster -and decision tree analysis
	X			Lin et al. (2011)	predicting a 4-level triage scale			cluster analysis, data mining
	X	X		Michalowski et al. (2007)	implementing a DSS on a handheld device			software engineering principles and expert consultation
		X		Davies (2007)	'see and treat' vs 'see' and 'treat'	X		
		X		Konrad et al. (2013)	split-flow process	X		
		X		Medeiros et al. (2008)	provider directed queueing	X		
			X	Dobson & Sainathan (2011)	cost of prioritization vs. benefits of prioritization		X	
			X	Alizamir et al. (2013)	accuracy vs. congestion		X	

Table 4: Overview of triage articles

priority such that the priority of a patient increases at a certain accumulation rate that will be higher for higher-urgency patients. When a server becomes idle, the patient with the highest accumulated priority is treated first. Sharif et al. (2014) provide an algorithm to find the most robust accumulation rates for each patient category that can achieve pre-specified key performance indicators.

Argon & Ziya (2009) suggest an original framework for prioritization in any general service setting. In their system, there are two types of 'customers' (high priority and low priority customers) and each customer provides a signal which is an imperfect indicator of the customer's priority. The higher the signal, the higher the probability that the customer is of the high priority type. Aiming to minimize the waiting cost, they compare a policy that always gives priority to the customer with the highest signal, policies with just 2 priority classes, and an extension of the generalized $c\mu$ rule of van Mieghem (1995) that takes into account both waiting cost and service times. An important result is that increasing the number of priority classes decreases the waiting costs in their setting.

Lastly, Ashour & Kremer (2013) assess the FAHP-MAUT triage algorithm through simulation. This algorithm first uses data on the chief complaint, vital signs, age, gender, and pain level as inputs for a fuzzy analytic hierarchy process (FAHP). Thereafter, multi-attribute utility theory (MAUT) is employed to prioritize patients based on their utility value.

Thanks to the digitization of patient records and the ability to more easily collect and store vast amounts of data, it is now also possible to uncover the hidden rules of triage. Table 4 lists some articles that use data mining, regression, and clustering techniques to quantify the relationship between patient characteristics (like age, gender, pain level, temperature, and heart rate) and the assigned triage scale or diagnosis. Chonde et al. (2013), Lin et al. (2010), and Lin et al. (2011) aim to classify patients according to the triage system currently used in their hospital. All three articles plan to use the obtained knowledge in expert systems or decision support systems to lower the cognitive stress and load on the triage nurse and to assist her in making better triage decisions. In that view, Michalowski et al. (2007) describe in detail how to design a proper human-friendly handheld device that implements the rule-based decision model and does not obstruct the medical tasks of the staff in any way.

Triage is usually performed by a *triage nurse* although physician-led triage and team triage are currently gaining attention (Burström et al., 2012; Oredsson et al., 2011; Welch & Savitz, 2012). Extending the tasks and responsibilities of the triage nurse with protocols that allow triage nurses to initiate diagnostic testing and treatments has proved to facilitate patient care in the ED and possibly even decrease LOS (Robinson, 2013). Davies (2007), Konrad et al. (2013), and Medeiros et al. (2008) evaluate alternative configurations for ED triage through simulation. Davies (2007) compares a 'see and treat' policy (patients with minor injuries are immediately treated by the same practitioner that assessed their condition) to a 'see' and 'treat' policy (the assessment step is performed by a highly skilled worker and separated from the treatment step that can be performed by a less-skilled nurse). He concludes that the latter policy utilizes the highly skilled resources more efficiently, which is more in line with lean principles. Konrad et al. (2013), on the other hand, consider a split-flow process where lower acuity patients (ESI 5, 4, or sometimes 3) are separated from the higher acuity patients (ESI 1, 2, or sometimes 3) in the ED. Triage in this system consists of an initial 'quick look' triage by a registered nurse and/or team triage, followed by a determination of the likelihood of admission. Medeiros et al. (2008), put a *provider directed queueing* system to the test. Here, an emergency care physician is placed at triage and works in a team to provide the resources necessary for the patient's care.

Finally, there are two articles that trade off several factors to determine whether prioritization is beneficial or not in a service system. While they discuss prioritization or diagnostics in a more general service system, they both mention that their analysis may also be relevant for ED triage. While most studies on prioritization assume that sorting is free and instantaneous, Dobson & Sainathan (2011) state that prioritization has both benefits and costs. The benefits are that more urgent patients are treated faster and that some information about the patient may already be obtained so that the following processes go more smoothly. The costs include the wage of the employees that perform the sorting and the time it takes to do the sorting. They investigate for which system parameters (e.g. the fraction of urgent patients and the difference in waiting costs between urgent and non-urgent patients) prioritization can decrease waiting or total costs. Furthermore, sorting is less attractive if misclassification errors are considered. Alizamir et al. (2013) focus on the trade-off between the accuracy of the diagnosis and the congestion in the system and model the problem as a partially observed Markov decision process. Whether additional tests, time, and effort is invested to improve the diagnosis accuracy depends on the number of tests that have already been performed and the prevailing level of congestion. Although these trade-offs can be very interesting, they may be less relevant in an ED context. Indeed, the accuracy of the prioritization is extremely important and the processing time of triage is usually negligible compared to the time the patient later spends waiting and undergoing treatment in the ED (Van der Vaart et al., 2011). Indeed, as Mullen (2003) highlights, the clinical priority of patients cannot be overlooked in the search for shorter waiting times or lower costs.

2.1.3 Queue management

In most service systems, customers have to wait to get access to the service system or even between different service steps; these waits are undesirable (Bitran et al., 2008). Several articles from medical journals indicate that high ED occupancy, long waits and consequently large LOS are the most important factors that influence LWBS rates (Fernandes et al., 1997; Hobbs et al., 2000). Evidence that other factors like demographics or day-of-week effects influence LWBS is far less consistent (Melton et al., 2014). Apart from customer balking (leaving immediately if the queue is too long [Kelton et al., 2010]) and reneging (joining the queue but leaving later, when the wait is taking too long [Kelton et al., 2010]), waiting may also affect the probability that customers return for service. Van Ackere et al. (2013) introduce this kind of feedback in service systems and derive guidelines for making capacity decisions using feedback diagrams and differential equations. In an ED setting, a similar kind of feedback may exist if patients can choose between different hospitals in a certain area; they will let their choice depend on previous experiences.

There are two possible ways to avoid the negative consequences of long waits. Firstly, an obvious solution is to reduce the waiting times themselves. In an ED setting, adding staff or beds, faster treatment procedures, or improving the flow through the ED are just some possible strategies to reduce waiting times. Secondly, good queue management could reduce the *perceived* waiting time (Katz et al., 1991), thereby also increasing patient satisfaction (Boudreaux & O'Hea, 2004; Thompson et al., 1996). Queue management is the focus of this section.

How to make waiting more pleasant (or rather bearable) is a reasonably well-covered topic in the marketing and psychology literature (Durrande-Moreau, 1999; Katz et al., 1991) and the insights from behavioral sciences are slowly finding their way to the OR field (Bitran et al., 2008; Larson, 1987; Nie, 2000). Batt & Terwiesch (2013) use regression models to reveal that the reneging behavior in EDs not only depends on the length of the wait; the observed queue length and the observed flow of patients in and out of the waiting room also influence the abandonments. They established that patients make assumptions about the severity of other patients' conditions and will respond differently to the arrival or departure of a (estimated) more or less severe patient. Currently, queueing models do not account for these kinds of behavior.

Most articles that are listed in Table 5 are inspired by call centers. While theoretical research on telephone queue models is quite extensive, research on actual patience patterns observed in practice is scarce. In EDs, on the other hand, some statistical research has been done on the LWBS patients (Batt & Terwiesch, 2013; Fernandes et al., 1997; Hobbs et al., 2000) but analytical models are lacking (Mandelbaum & Zeltyn, 2013). There are some important differences between call center settings and ED settings; the number of servers tends to be larger and the willingness to wait tends to be lower in call centers. However, they also have important similarities; both settings typically have time-varying arrivals and while the queue in an ED may not be completely invisible like in a call center, triage and priority rules make it very hard for patients to accurately estimate their own expected delay. These similarities make articles on abandonments in call centers relevant for ED research, especially since they fill the apparent gap in the literature on analytical models for abandonments in the ED. Section 2.1.3.1 discusses two articles that recommend new ways to model balking and reneging, based on the fact that they may not only be influenced by the actual waiting but also by how this waiting time is being perceived. Section 2.1.3.2 focuses on nine articles that evaluate and compare the accuracy of different types of predictions of the customers' waiting time. The predictions could be used to inform patients about their expected delay. The last four articles of Table 5, finally, are reviewed in Section 2.1.3.3. They consider the customer and service provider to be strategic players in a 'game' where service providers try to influence customers by (possibly vague) waiting time announcements and customers respond to these announcements by deciding to stay in the queue, balk or renege with the aim to optimize their own utility (weighing waiting costs against the value of the service).

Research goal	Article	Methodology		Balking	Reneging	Queueing discipline	Announcement type	endogenous customer reactions
		Simulation	Queueing					
modeling balking and reneging	Mandelbaum & Zeltyn (2013)	X	M/M/n +M	X	X	/	/	/
	Akşin et al. (2013)	X	X	X	X	/	/	/
evaluating announcement types	Ibrahim & Whitt (2011b)	X	M(t)/M/s+GI		X	FCFS	real-time estimators	
	Ibrahim & Whitt (2009a)	X	GI/M/s			FCFS	real-time estimators	
	Ibrahim & Whitt (2009b)	X	GI/GI/s+GI		X	FCFS	real-time estimators	
	Ibrahim & Whitt (2011a)	X	M(t)/GI/s(+GI)		X	FCFS	real-time estimators	
	Ibrahim et al. (2014)	X	M/M/N +M	X	X	FCFS	real-time estimators	X
	Armony et al. (2009)		G/GI/s+GI	X	X	FCFS	real-time estimators	X
	Jouini et al. (2014)	X	M(t)/M/s(t)	X	X	priority	derived from distribution	
	Jouini et al. (2011)		M/M/s+M	X	X	FCFS	derived from distribution	X
	Jouini et al. (2009)		M/M/s+M	X	X	priority	derived from distribution	X
strategic customers and service providers	Allon & Bassamboo (2011)		M/M/N	X	X	FCFS	vague and non-quantitative	X
	Allon et al. (2011)		M/M/1	X	X	FCFS	system state / vague & non-quantitative	X
	Guo & Zipkin (2007)		M/M/1	X		FCFS	system state/exact waiting time	X
	Plambeck & Wang (2013)		M/M/1	X		FCFS	system state/exact waiting time	X
	Shone et al. (2013)		M/M/1	X		FCFS	queue length	X

Table 5: Overview of articles on queue management

2.1.3.1 Modeling balking and reneging

Akşin et al. (2013) and Mandelbaum & Zeltyn (2013) express the need to understand and model balking and reneging behavior correctly when queueing models are used to approximate real-life service systems. Mandelbaum & Zeltyn (2013) introduce a framework for understanding impatience, distinguishing the times that a customer expects to wait, is required to wait, is willing to wait, actually waits and felt waiting. Akşin et al. (2013), on the other hand, model the customer's balking and reneging behavior endogenously instead of using an exogenous patience time distribution. Customers are assumed to make a wait or quit decision after each waiting period based on forward-looking behavior and rational utility maximization. They advocate that modeling the customers' behavior endogenously is important when major changes are to be implemented in the system.

2.1.3.2 Making accurate delay predictions

The philosophy behind making delay announcements is that an uncertain wait causes more stress and is perceived to take longer than certain waits (Durrande-Moreau, 1999). In an ED, these kinds of announcements may make the wait more bearable and deter patients from LWBS (Ibrahim & Whitt, 2011b; Sun et al., 2012). All nine OR articles that look into different types of predictions of the waiting time, use queueing theory and most of them also employ simulation to evaluate the efficiency of the prediction, often in a more realistic setting than what is possible in queueing theory. The articles differ in the system characteristics they consider and in the type of announcement that they evaluate. Three of them deal with time-varying arrivals (Ibrahim & Whitt, 2011a,b; Jouini et al., 2014) and two of them study priority queueing (Jouini et al., 2009, 2014). Since EDs have time-varying arrivals and patients with different acuities, these articles are of special interest. The last column of Table 5 indicates whether the customer reactions to the delay announcements are endogenized in the analysis or not. Although endogenizing the customer behavior makes the analysis more complex, it also makes it more realistic. If customers are provided an estimate of their waiting time, they may alter their balking and reneging behavior. Indeed, if the announced delay is longer than the customer is willing to wait, he may balk immediately. Furthermore, some customers may renege if they are waiting longer in reality than what was announced.

Most articles evaluate announcements that are based on real time estimators like delay-history-based and queue-length-based estimators. Delay-history-based estimators include; the delay of the last customer to enter service

(LES), the delay experienced so far by the customer at the head of the line (HOL), and the delay experienced by the customer that arrived most recently among those who have already completed service (RCS). The practical appeal of delay-history based estimators is that they need very little information about the system itself (Ibrahim & Whitt, 2009a, 2011a). These estimators can be accurate but may need refinements if arrivals are time-varying (Ibrahim & Whitt, 2009a), in case there are abandonments (Ibrahim & Whitt, 2009b), or if customer reactions to the announcements are endogenized (Ibrahim et al., 2014).

Jouini et al. (2014) propose two different delay estimators; one based on an Erlang distribution and one based on a Normal distribution. Using a newsvendor-like objective function allowing for different penalty costs for under- and overestimations, they concluded that the Erlang approximation gave the best results. Jouini et al. (2009, 2011), on the other hand, first fix a *delay coverage*, β, and consequently announce an expected waiting time, x, so that the probability of an actual waiting time smaller than x is equal to β.

2.1.3.3 Dynamics between service provider and customer

The last five articles of Table 5 assume that both the customers and the service provider are strategic players. Service providers will try to influence the customers' balking and reneging behavior using delay announcements. Customers, on the other hand, will try to maximize their own utility which is a function of waiting costs and the value they attribute to receiving the service. Allon & Bassamboo (2011) and Allon et al. (2011) model this strategic game as a Markov perfect Bayesian Nash equilibrium and the announcement they consider may be intentionally vague to lure customers. Guo & Zipkin (2007) compare three levels of information; no information, information about the system occupancy, or information about the exact waiting time. Customers decide to balk or stay based on the information they obtain. Both Guo & Zipkin (2007) and Shone et al. (2013) establish that whether providing more information is beneficial or not, for either the service provider or the customer, will depend on the system characteristics. Similarly, Plambeck & Wang (2013) argue that hospitals should not reveal the ED queue length or waiting time since it might deter patients who really do need treatment to enter the ED. Using *hyperbolic discounting*, they model patient preferences as if patients perceive the ED visit as unpleasant and lack self-control to undergo unpleasant services to achieve benefits (i.e. good health) in the long-run.

2.2 Discharge

The back end of the ED is the least studied area for improvement in OR literature (Saghafian et al., 2014a; Welch, 2012). This is surprising since inpatient boarding, a problem that is also situated at the back end of the ED, is often referred to as one of the most important causes of ED crowding. After patients finish treatment in the ED, they are either discharged home or need to be admitted to the hospital. While discharging patients home may be hampered by paperwork and rigid discharge procedures, or the inability or unwillingness of nursing homes or family members to pick up the patient (Moskop et al., 2009a), the major problem in most EDs lies with the patients that need to be admitted to an inpatient unit (IU). The inability to swiftly transfer care from the ED to the IU often forces these patients to stay in the ED. This phenomenon where patients wait in the ED to be admitted to the IU is called *'inpatient boarding'* (Chalfin et al., 2007; Moskop et al., 2009a; Pines et al., 2011b; Richardson & Mountain, 2009), *'access block'* (Crawford et al., 2013; Fatovich et al., 2005; Gilligan et al., 2008; Khanna et al., 2012; Luo et al., 2013; Richardson & Mountain, 2009) or *'bed block'* (Bair et al., 2010; El-Darzi et al., 1998; Helm et al., 2011) and has often been identified as the most important cause of ED crowding (Fatovich et al., 2005; Pines et al., 2011b; Rabin et al., 2012; Richardson & Mountain, 2009; Steele & Kiss, 2008; Trzeciak & Rivers, 2003). On top of that, inpatient boarding has also been associated with increased LWBS (Bernstein et al., 2009; Patel et al., 2014; Wiler et al., 2013), increased AD (Borders et al., 2009; Fatovich et al., 2005; Patel et al., 2014; Trzeciak & Rivers, 2003), worse patient outcomes and higher mortality rates (Bernstein et al., 2009; Chalfin et al., 2007; Singer et al., 2011; Sprivulis et al., 2006; Trzeciak & Rivers, 2003), decreased patient satisfaction with both the ED and the hospital in general (Pines et al., 2008), frustration among medical staff (Olshaker & Rathlev, 2006), higher ED LOS (Bernstein et al., 2009; Patel et al., 2014), and loss of revenue (Falvo et al., 2007; Pines et al., 2011a).

What makes the back end of the ED more challenging, is that the efficiency with which patients can be discharged is not entirely in the control of the ED. Although a lack of inpatient beds is mostly referred to as the number one cause of boarding, Armony et al. (2011) and Pines et al. (2011a) note that part of the ED back end problem may be caused by other inefficiencies in the system, needing more sophisticated solutions than just adding inpatient beds. Armony et al. (2011) classify the causes of delays in the transfer from ED to IU into 4 categories; ED-IU synchronization issues, bad work methods, a lack of staff availability, and a lack of equipment availability. Shi et al. (2014), on the other hand, distinguish two types of waiting due to secondary bottlenecks between the ED and the IU; pre-bed allocation delays and post-bed allocation delays. Depending on which process or resource causes the bottleneck at the back-end of the ED, the appropriate strategy may differ. Therefor, a first analysis of the congestion source and frequency as in Kolb et al. (2007) or Osorio & Bierlaire (2009) is inevitable.

We will introduce three types of improvement strategies, distinguished by which department is mainly responsible for the implementation of the strategy; IU solutions (Section 2.2.1), ED solutions (Section 2.2.2), and solutions involving a more collaborative effort encompassing several hospital departments (Section 2.2.3). Since the back end of the ED is still a fairly unexplored area in the OR/OM literature, some solutions that will be discussed here are only based on empirical evidence and have not been analyzed from a more theoretical OR/OM perspective. However, they may provide a good foundation for future research.

2.2.1 IU solutions

Since a lack of *inpatient beds* is most often mentioned as the cause of delays in the transfer of patients from the ED to the IU, adding more staffed IU beds seems the most obvious solution (Olshaker & Rathlev, 2006; Pines et al., 2011b). A common rule-of-thumb to determine the appropriate number of beds in each IU is to strive for a target utilization of 85% (Green, 2002). However, several articles point out that this target is too simple and often suboptimal since the size of the department and the variation in the IU demand and LOS will influence the performance of the target (de Bruin et al., 2007, 2010; Khanna et al., 2012; Kuntz et al., 2014; Luangkesorn et al., 2012). The specific system characteristics of each IU should be taken into account when deciding on a target since setting an erroneous target utilization can have dangerous consequences; Kuntz et al. (2014) use discrete-time survival analysis to show that hospitals experience safety tipping points as the utilization increases and that operating at a utilization above these safety tipping points can substantially increase in-hospital mortality. The existence of these safety tipping points may call for more flexible beds. In that spirit, van Essen et al. (2013) propose to cluster units so that the probability of not being able to admit a patient is acceptably small. Litvak et al. (2008) even aim to reserve a small number of ICU beds in several hospitals in a certain region to accommodate emergency patients. By sharing these beds with all hospitals in the same region, less beds are needed in total. As Table 6 shows, simulation is often applied to improve the allocation of beds to IUs. All simulation articles choose DES, except for Vanderby & Carter (2010) who employ SD to model the hospital patient flow. Queueing models are also gaining popularity and are evolving into more and more complex models that are able to incorporate the blocking behavior, multiple wards, and arbitrary patient routings. An alternative way to increase the capacity of the IU without adding beds is to use the IU hallways. Exploiting IU hallways is often applied in practice and extensive surveys show that patients usually prefer waiting in the IU over waiting in the ED (Garson et al., 2008; Walsh et al., 2008). Furthermore, letting patients board in the ED does not only negatively affect their satisfaction with the ED but also with the hospital in general; moving waiting patients from the ED to the IU may thus significantly improve patient satisfaction (Pines et al., 2008).

Another area for improvement is the *discharge process* of the IU. Firstly, we introduce strategies that make the discharge process of the IU more in line with the emergency arrivals to the IU. It is common practice for physicians to make just one round per day, in the afternoon, in which they assess the IU patients and make discharge decisions. This practice tends to result in a build-up of boarding patients right before the IU discharge round. Sometimes, starting the discharge round earlier in the day so that the IUs are emptied before the bulk of new emergency patients arrive can significantly reduce boarding (Ferrin et al., 2007; Khanna et al., 2011, 2012; Powell et al., 2012; Shi et al., 2014; Vermeulen et al., 2009). Predictions for the arrival of emergency patients to the IU over the day (Cochran & Roche, 2008; Gallivan & Utley, 2011; Peck et al., 2012, 2013) can be useful to gain insights in the proper timing of the discharge rounds.
Physicians naturally tend to discharge patients earlier when the IU is busier (Anderson et al., 2011; Forster et al., 2003). Hosseinifard et al. (2014) call this kind of early discharges from the ICU as a reaction to access block 'bumping', 'demand-driven discharge', or 'premature discharging'. Since discharging patients earlier may also increase the probability of readmission or negative patient outcomes (Chan et al., 2012a; Kc & Terwiesch, 2009, 2012), it is important to make a balanced trade-off between the advantage of reducing the current occupancy and the possible negative effects for the patients of an early discharge. Some articles propose dynamic discharge policies that account for both the occupancy of the hospital and the medical status of the patient in their discharge policies (Berk & Moinzadeh, 1998; Chan et al., 2012a; Crawford et al., 2014). 'Reverse triage' systems help to decide on which patient to discharge first, by classifying patients according to 'readiness to depart' (Moskop et al., 2009b). In case of a sudden surge in demand, the reverse triage class of the patients, which should be readily

available, can guide the decision of which patient to quickly send home (Kelen et al., 2001, 2006). Hosseinifard et al. (2014) use dynamic programming and simulation-optimization to show that both the readmission risk and the remaining LOS of the patients should be taken into account when making discharge decisions. Instead of discharging patients, Thompson et al. (2009) suggest proactive reallocation of patients to less crowded IUs before the bulk of new patients comes in from the ED. Secondly, since requests for admission of ED patients can occur at any time of the day, smoothing the discharge process of the IU over time also proves advantageous. The variability in the discharge process can, for instance, be reduced by making twice-daily rounds instead of just one discharge round each day (Howell et al., 2008, 2010). Wong et al. (2010) use SD simulation to show that smoothing the discharge pattern over the week, avoiding differences between weekdays and weekends, can also help.

Finally, the *scheduling of elective patients* can be improved. While the arrival of ED patients is random, elective patients can be scheduled in advance. Since several studies show that the variability in these scheduled admissions is often larger than the variability in emergency arrivals, it is clear that there is room for improvement (Bekker & Koeleman, 2011; Luo et al., 2013; Marsh et al., 2004; McManus et al., 2003). For instance, the common practice to admit many elective patients on Monday will significantly increase boarding in the beginning of the week. When scheduling elective patients, the expected emergency admissions can already be taken into account (Adan et al., 2011; Bachouch et al., 2012; Ceschia & Schaerf, 2011; Helm & Van Oyen, 2014; Vissers et al., 2007). Helm et al. (2011), as already discussed in Section 2.1.1, propose a new 3^{rd} gateway for patients that do not have a very urgent condition but who cannot wait too long to get treatment. These patients will be called in whenever the hospital utilization is low so that hospital utilization will be smoothed over time and the ED load is reduced.

Goal	Article	Methodology: Simulation	Queueing	Mathematical programming
Optimizing IU bed capacity	Ferrin et al. (2007); Harper & Shahani (2002); Luangkesorn et al. (2012); Mustafee et al. (2012); Vanderby & Carter (2010)	X		
	Cochran & Bharti (2006); Cochran & Roche (2008); Koizumi et al. (2005); Patrick (2011)	X	X	
	de Bruin et al. (2007, 2010); Lin et al. (2013); Litvak et al. (2008); Shi et al. (2014); van Essen et al. (2013)		X	
	Bretthauer et al. (2011)		X	X
	Kuntz et al. (2014)	discrete-time survival analysis		
Optimizing the discharge process	Crawford et al. (2014); Ferrin et al. (2007); Wong et al. (2010)	X		
	Thompson et al. (2009)	X	X	
	Hosseinifard et al. (2014)	X		X
	Berk & Moinzadeh (1998); Chan et al. (2012a); Shi et al. (2014)		X	
	Kc & Terwiesch (2009, 2012)	econometric analysis		
Optimizing elective patient scheduling	Vissers et al. (2007)	X		
	Adan et al. (2011)	X		X
	Helm et al. (2011)		X	
	Bekker & Koeleman (2011); Helm & Van Oyen (2014)		X	X
	Bachouch et al. (2012); Ceschia & Schaerf (2011)			X

Table 6: Overview of articles that explore IU solutions for ED back-end problems

2.2.2 ED solutions

The process of discharging patients home can be hindered because of administrative red tape, or patients may have to wait for prescriptions, medication, instructions for their future treatment, or transportation. Providing a separate area for these waiting patients, a *discharge lounge or kiosk*, is highly recommended (Moskop et al., 2009b; Rabin et al., 2012; Welch, 2012). The potential of a discharge lounge has been assessed using simulation by Alavi-Moghaddam et al. (2012), Ferrin et al. (2007), and Kolb et al. (2008).

A *holding area*, sometimes called an express admission unit (Welch, 2012), is similar to a discharge lounge but intended for patients that are waiting for an IU bed. In contrast with patients that can be discharged home, patients awaiting admission might not be medically stable and need care while they are waiting (Armony et al., 2011). Kolb et al. (2008) use simulation to show that a holding area can improve the patient flow in the ED, especially if a large share of the emergency patients need to be admitted.

An *observation unit* (Armony et al., 2011; Borders et al., 2009; Crawford et al., 2013; Derlet & Richards, 2000; Pines et al., 2011b) or *clinical decision unit* (Moskop et al., 2009b; Welch, 2012) differs from the holding area described above in the sense that it offers a place to stay for patients that will likely go home but need to stay in the hospital for an extended time (but usually less than 24 hours). It accommodates patients that do not really need admission but are not well enough to immediately go home. Some patients need to stay for observation, making sure their condition does not worsen, others need additional tests to make sure the diagnosis is correct. While Moloney et al. (2006) describe a successful implementation in practice, Kolb et al. (2008) investigate its potential using DES. Lovejoy & Desmond (2011) apply Little's law and 'real options' to get an estimate of the needed size of an observation unit.

Finally, there are several studies that aim to optimize the *resource capacity* (staff and beds) in the ED, taking into account the blocking that may result from a lack of capacity in the subsequent IU. Both simulation (Ferrin et al., 2007) and queueing theory (de Bruin et al., 2010; Lin et al., 2013) are applied. Khare et al. (2009) show, using DES, that adding ED beds is not beneficial at all if these beds are not a bottleneck. Their simulation model illustrated an interesting analogy; if you compare the ED with a pipeline, adding ED beds will increase the width of the pipe in the middle but if the width of the pipe at the end is not increased, throughput will not increase and, even worse, LOS may even increase.

Strategy	Article	Methodology	
		Simulation	Queueing
discharge lounge	Alavi-Moghaddam et al. (2012); Ferrin et al. (2007); Kolb et al. (2008)	X	
holding area	Kolb et al. (2008)	X	
observation unit	Kolb et al. (2008)	X	
adding capacity	Ferrin et al. (2007)	X	
	de Bruin et al. (2010); Lin et al. (2013)		X

Table 7: Overview of articles that explore ED solutions for ED back-end problems

2.2.3 Collaborative solutions

More and more researchers are striving for 'collaborative' solutions meaning that they require the involvement of more than one hospital department. Here, we will focus on projects that require some commitment of both the ED and the IU to be successful. This kind of solutions may be harder to implement but worth the effort (see also Armony et al., 2011; Drupsteen et al., 2013; Moskop et al., 2009a; Patel et al., 2014). Typically, though, these collaborative efforts include 'softer' approaches which are harder to analyze and evaluate using OR/OM techniques. Some examples are clearer responsibilities and goals, leadership programs, and new staff or multidisciplinary teams for more coordinated, active bed management (Borders et al., 2009; Hemphill & Nole, 2005; Howell et al., 2008, 2010; Marsh et al., 2004; Patel et al., 2014; Trzeciak & Rivers, 2003).

An important step in collaboration is sharing information. Real-time hospital-wide information, in particular, has proven to be advantageous in practice (Hemphill & Nole, 2005; Marsh et al., 2004; Patel et al., 2014). Peck et al. (2012, 2013), for example, use real-time information on the ED to predict admissions to the IU in the near future. These kinds of predictions can subsequently be used by a 'bed management team' to proactively discharge patients (see Section 2.2.1) or in *routing policies* that regulate which ED patient to send to which unit. Indeed, while most patients may have a preferred destination IU, allowing for some flexibility in the IU assignment can decrease the total necessary capacity of the IUs (Pines et al., 2011b). Although this kind of dynamic admission decisions has been studied in general production settings (e.g. Pines et al., 2011b), optimizing the assignment of patients to IUs, 'routing policies', have been scarcely studied in the OR/OM literature as shown in Table 8. Cohen et al. (1980) compare how either blocking a patient or assigning them to an alternative unit (according to a predefined preference list) will affect the necessary capacity of the IU in a DES study. The other two articles both employ queueing theory. Mandelbaum & Zeltyn (2013) compare three routing policies with respect to fairness (from the IU staff's point of view), ease of implementation, and efficiency. Shi et al. (2014) include a simple threshold policy in their model to obtain realistic results; when the boarding time of a patient exceeds a time-varying threshold, he will be sent to an alternative IU instead of waiting any longer to get into his preferred IU.

Strategy	Article	Methodology	
		Simulation	Queueing
ED - IU routing policies	Cohen et al. (1980)	X	
	Mandelbaum & Zeltyn (2013); Shi et al. (2014)		X

Table 8: Overview of articles that explore collaborative solutions for ED back-end problems

3 Doctoral proposal

Based on the literature and a real-life case study, I have distilled the conceptual model of Figure 2 as a realistic representation of an ED. It can be seen that there are three phases where patients wait; (1) waiting for a bed, (2) waiting for staff to perform the next treatment, or (3) waiting for admission in the IU (boarding). It should be clear that, the longer patients wait for either staff or admission, the longer newly arriving patients will have to wait for a bed (since bed occupancy will rise if either (2) or (3) increases).

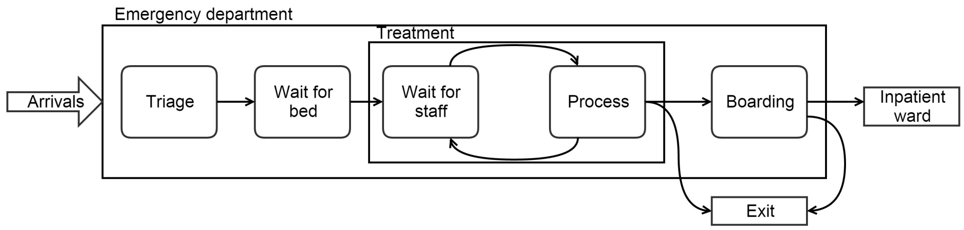

Figure 2: Conceptual model of patient flow in the ED

Given the severity of the inpatient boarding problem in practice and the lack of OR research on this topic, I intend to explore this problem and use analytical methods to develop improvement strategies. I will first highlight two queueing systems which will be the starting point of my research; Section 3.1 elaborates on a very basic queueing network and Section 3.2 will present a slightly more refined queueing system. Section 3.3 describes a final research topic; collaboration of resources. Lastly, Section 3.4 discusses the time frame for each of the research topics. Note that both the research topics and the time frame are flexible. If new opportunities arise or the planned research topics turn out not to be as interesting or feasible, we may redirect the research topics and/or adjust the schedule.

3.1 The simple queueing model

Inpatient boarding will be incorporated in the model as an additional waiting process where the patient does not demand treatment but is still occupying one of the ED beds. The goal is to start with a simple system (Figure 3) that can be expanded and refined over time. Notice the 2 queues in Figure 3; they reflect the fact that a patient will first have to seize a bed before he can wait for staff to start treatment. The ED has a total of N beds and s_t servers (doctors or nurses; for now we assume there is just one type of staff), $t \geq 0$. Notice that, while the number of servers in the ED may change over time, the number of beds is assumed to be fixed. Patients arrive according to a time-inhomogeneous Poisson process with rate function λ_t, $t \geq 0$. Assuming homogeneous patients and a FCFS queueing policy, the patients must first undergo treatment with independent and identically distributed (i.i.d) exponential service times with mean $1/\mu_a$. After treatment, patients must undergo an additional boarding step with probability p_b and exponential service times with mean $1/\mu_b$. After the boarding process, the patient leaves the system either through discharge or through admission to an IU. The ED bed now becomes available again for a new patient.

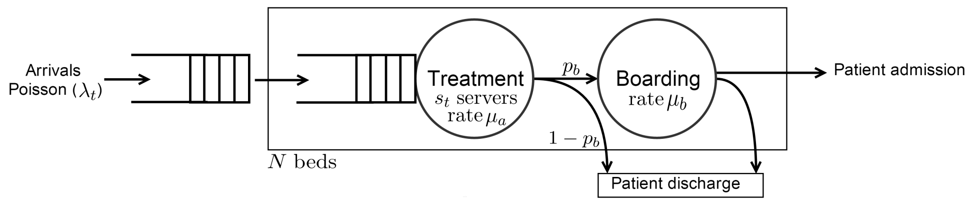

Figure 3: A diagram of the simple queueing network

The most interesting part of this queueing model is the fact that the number of patients that can simultaneously receive treatment, Q_t, is bound by either the number of available servers or the number of beds that are available for patients needing treatment, the *non-blocked beds* (which will depend on the number of boarding patients). Suppose the number of boarding patients at time t is equal to B_t. The number of non-blocked beds will be $N - B_t$. Q_t will then have to be smaller or equal to $min\{s_t, N - B_t\}$. When optimizing resource requirements, beds and staff may have to be taken into account simultaneously. It is expected that higher p_b and lower μ_b will require more beds to maintain low waiting times and LOS but it is yet unclear how these parameters might influence staffing levels. Furthermore, this model inherently contains an interesting combination of decision making on the strategic level (determining the necessary number of beds in an ED or the number of beds in the IU that determines the boarding times is typically a strategic decision) and the tactical level (scheduling staff) (Zeltyn et al., 2011). Possible extensions to this model are;

- **non-homogeneous patients**: Allowing for non-homogenous patients opens several doors to make the model more realistic (and complicated). Firstly, the queueing discipline may be changed to priority queueing as is common in practice (see Section 2.1.2 on triage in EDs). Secondly, there seems to be a trend to segment patients into streams, such as fast tracks, for more efficient health care delivery (Welch, 2009). Each of these patient streams may have a different priority, mean service time, admission probability and possibly even boarding time (boarding times will be longer for patients that need to go to inpatient units that typically have high utilizations). An interesting research topic is to investigate how sharing beds and staff or dedicating them to the separate patient stream will affect patient flow and resource requirements.

- **generally distributed service times**: From a queueing theory perspective, it is convenient to start the analysis with exponentially distributed service times. However, it has been shown that this may not be the most realistic service distribution (Holm & Barra, 2011). Some recent articles even go so far as to propose endogenous service times in healthcare services (Anand et al., 2011; Armony et al., 2011; Batt & Terwiesch, 2012; Chan et al., 2012b; Kc & Terwiesch, 2009)

- **LWBS**: Patients that leave without being seen are a problem in emergency departments (see Section 2.1.3 for more references). Incorporating balking in a queueing model may lead to very different behavior and conclusions (Batt & Terwiesch, 2013). An important consideration in modeling the balking behavior, is that the probability of balking is likely to be lower for patients waiting for staff in a bed than for patients who are still waiting for a bed in the queue.

3.2 Queueing model with reentrant patients

After gaining insights and experience from the simple queueing model, I will proceed by incorporating the Erlang-R queueing model of Yom-Tov & Mandelbaum (2014) into the 'treatment' step of the previous model. The 'R' stands for reentrant customers; it can be applied in the emergency setting by taking the point of view of one type of staff, the physicians for instance. Patients will be treated by a physician, then they may have to receive treatment by a nurse, or undergo tests and scans after which the physician may visit the same patient again. In that way, patients return to service with the physician (reenter) several times during their ED visit. After an initial generally distributed process step with mean $1/\mu_a$, the patient may return for additional treatment by the physician with a probability p after a delay that is generally distributed with a mean of $1/\delta$.

The Erlang-R queueing model should allow for more realistic modeling of the treatment process in the ED and consequently better resulting staffing levels to accommodate the time-varying arrivals. Inspiration for further refinements of the system, in addition to the extensions that have been suggested for the simple model, are;

- **priority queueing**: Erlang-R assumes FCFS queueing. It may be interesting to find out if it is useful to give priority to patients that have already received a first service or patients that are still waiting in queue. Especially if this is combined with LWBS rates that are larger for patients waiting in queue than for patients that already received a first treatment

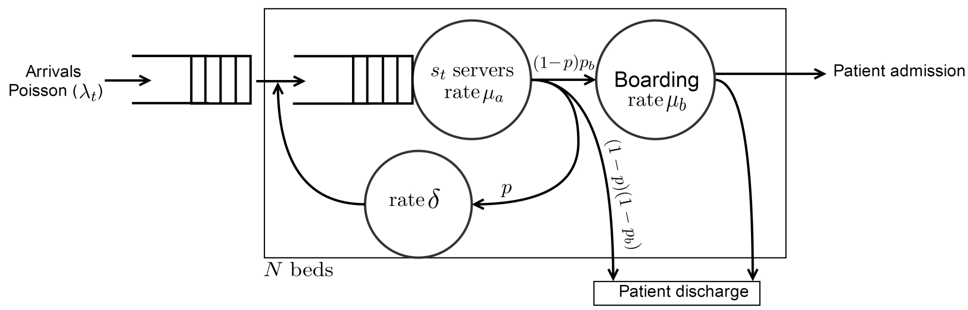

Figure 4: A diagram of the refined queueing network. Erlang-R is used to model the treatment in the ED

- **multitasking and case managers**: The Erlang-R model assumes that patients can be treated by any of the physicians. In practice, however, patients will usually receive all treatments from the same physician that gave them their first treatment. Campello et al. (2013) call these servers that are assigned to multiple customers and have frequent, repeated interactions with each customer until the customer's service is completed '*case managers*'. Simulation would allow to assess how this decrease in flexibility might affect the performance. Recent articles report on the consequences of multitasking on the quality and speed of the ED treatment (Kc, 2014) and how to determine an appropriate number of patients per physician (Campello et al., 2013).

3.3 Collaboration of resources

Gurvich & Van Mieghem (2014) bring up the interesting topic of collaboration in networks and how this may affect bottleneck analysis, capacity of the network, and ultimately throughput. In EDs, nurses and physicians sometimes need to assist each other. Gurvich & Van Mieghem (2014) show that this kind of simultaneous collaboration of multiple human resources requires synchronization of the resources and may lead to *unavoidable bottleneck idleness (UBI)* which, in turn, will lower the maximum achievable throughput or actual capacity of the system. Since Gurvich & Van Mieghem (2014) focus on small 'toy problems', the goal is to get a feeling for how collaboration affects the performance of a realistic ED network and to develop a heuristic that enhances the staffing levels, taking the consequences of collaboration on throughput into account. One article that considers collaboration in EDs is Lim et al. (2013). They enhance a simulation model by explicitly accounting for the interaction between physicians and their delegates and discuss the consequences for the simulation results.

3.4 Time frame

I have 34 months left to finish my PhD and propose the schedule of Figure 5 for the coming years. I plan to spend 9 months on each of the first two research topics and 6 months on the collaboration heuristic. This leaves some time for a proper analysis of the research results, compiling the results into scientific articles, and reporting on the results in internal doctoral seminars. I also intend to present the work in progress or results in international conferences to get feedback and discuss with other researchers working on similar topics (not displayed on Figure 5).

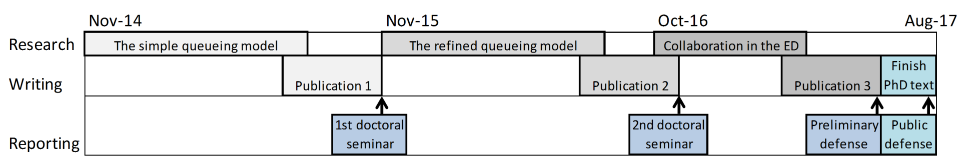

Figure 5: A timeline for the different research topics

References

Adan, I., Bekkers, J., Dellaert, N., Jeunet, J., & Vissers, J. (2011). Improving operational effectiveness of tactical master plans for emergency and elective patients under stochastic demand and capacitated resources. *European Journal of Operational Research*, *213*(1), 290 – 308.

Akşin, Z., Ata, B., Emadi, S. M., & Su, C.-L. (2013). Structural estimation of callers' delay sensitivity in call centers. *Management Science*, *59*(12), 2727 – 2746.
URL `http://pubsonline.informs.org/doi/abs/10.1287/mnsc.2013.1730`

Alavi-Moghaddam, M., Forouzanfar, R., Alamdari, S., Shahrami, A., Kariman, H., Amini, A., Pourbabaee, S., & Shirvani, A. (2012). Application of queuing analytic theory to decrease waiting times in emergency department: Does it make sense? *Archives of Trauma Research*, *1*(3), 101.

Alizamir, S., de Véricourt, F., & Sun, P. (2013). Diagnostic accuracy under congestion. *Management Science*, *59*(1), 157 – 171.

Allon, G., & Bassamboo, A. (2011). The impact of delaying the delay announcements. *Operations Research*, *59*(5), 1198 – 1210.

Allon, G., Bassamboo, A., & Gurvich, I. (2011). "We will be right with you": Managing customer expectations with vague promises and cheap talk. *Operations Research*, *59*(6), 1382 – 1394.

Allon, G., Deo, S., & Lin, W. (2013). The impact of size and occupancy of hospital on the extent of ambulance diversion: Theory and evidence. *Operations Research*, *61*(3), 544 – 562.
URL `http://pubsonline.informs.org/doi/abs/10.1287/opre.2013.1176`

Almehdawe, E., Jewkes, B., & He, Q.-M. (2013). A Markovian queueing model for ambulance offload delays. *European Journal of Operational Research*, *226*(3), 602 – 614.
URL `http://www.sciencedirect.com/science/article/pii/S037722171200882X`

Anand, K. S., Paç, M. F., & Veeraraghavan, S. (2011). Quality–speed conundrum: Tradeoffs in customer-intensive services. *Management Science*, *57*(1), 40 – 56.

Anderson, D., Price, C., Golden, B., Jank, W., & Wasil, E. (2011). Examining the discharge practices of surgeons at a large medical center. *Health Care Management Science*, *14*(4), 338 – 347.

Argon, N. T., & Ziya, S. (2009). Priority assignment under imperfect information on customer type identities. *Manufacturing & Service Operations Management*, *11*(4), 674 – 693.
URL `http://dx.doi.org/10.1287/msom.1080.0246`

Armony, M., Israelit, S., Mandelbaum, A., Marmor, Y. N., Tseytlin, Y., & Yom-Tov, G. B. (2011). Patient flow in hospitals: A data-based queueing-science perspective. *Submitted to Stochastic Systems*, *20*.

Armony, M., Shimkin, N., & Whitt, W. (2009). The impact of delay announcements in many-server queues with abandonment. *Operations Research*, *57*(1), 66 – 81.

Asamoah, O. K., Weiss, S. J., Ernst, A. A., Richards, M., & Sklar, D. P. (2008). A novel diversion protocol dramatically reduces diversion hours. *The American Journal of Emergency Medicine*, *26*(6), 670 – 675.

Ashour, O. M., & Kremer, G. E. O. (2013). A simulation analysis of the impact of FAHP-MAUT triage algorithm on the emergency department performance measures. *Expert Systems with Applications*, *40*(1), 177 – 187.
URL `http://www.sciencedirect.com/science/article/pii/S0957417412008792`

Asplin, B. R., Magid, D. J., Rhodes, K. V., Solberg, L. I., Lurie, N., & Jr., C. A. C. (2003). A conceptual model of emergency department crowding. *Annals of Emergency Medicine*, *42*(2), 173 – 180.
URL `http://www.sciencedirect.com/science/article/pii/S019606440300444X`

Au, L., Byrnes, G., Bain, C., Fackrell, M., Brand, C., Campbell, D., & Taylor, P. (2009). Predicting overflow in an emergency department. *IMA Journal of Management Mathematics*, *20*(1), 39 – 49.

Bachouch, R. B., Guinet, A., & Hajri-Gabouj, S. (2012). An integer linear model for hospital bed planning. *International Journal of Production Economics*, *140*(2), 833 – 843.
URL `http://www.sciencedirect.com/science/article/pii/S0925527312003325`

Bair, A., Song, W., Chen, Y.-C., & Morris, B. (2010). The impact of inpatient boarding on ED efficiency: A discrete-event simulation study. *Journal of Medical Systems*, *34*(5), 919 – 929.
URL `http://dx.doi.org/10.1007/s10916-009-9307-4`

Batt, R. J., & Terwiesch, C. (2012). Doctors under load: An empirical study of state-dependent service times in emergency care. Working paper, University of Wisconsin-Madison, Madison.

Batt, R. J., & Terwiesch, C. (2013). Waiting patiently: An empirical study of queue abandonment in an emergency department. The Wharton School, University of Pennsylvania, Philadelphia, PA.

Bekker, R., & Koeleman, P. M. (2011). Scheduling admissions and reducing variability in bed demand. *Health Care Management Science*, *14*(3), 237 – 249.

Berk, E., & Moinzadeh, K. (1998). The impact of discharge decisions on health care quality. *Management Science*, *44*(3), 400 – 415.

Bernstein, S. L., Aronsky, D., Duseja, R., Epstein, S., Handel, D., Hwang, U., McCarthy, M., John McConnell, K., Pines, J. M., Rathlev, N., Schafermeyer, R., Zwemer, F., Schull, M., Asplin, B. R., & Society for Academic Emergency Medicine, E. D. C. T. F. (2009). The effect of emergency department crowding on clinically oriented outcomes. *Academic Emergency Medicine*, *16*(1), 1 – 10.
URL `http://dx.doi.org/10.1111/j.1553-2712.2008.00295.x`

Bhattacharjee, P., & Ray, P. K. (2014). Patient flow modelling and performance analysis of healthcare delivery processes in hospitals: A review and reflections. *Computers & Industrial Engineering*, (0), –.
URL `http://www.sciencedirect.com/science/article/pii/S0360835214001272`

Bitran, G. R., Ferrer, J.-C., & Rocha e Oliveira, P. (2008). OM forum-managing customer experiences: Perspectives on the temporal aspects of service encounters. *Manufacturing & Service Operations Management*, *10*(1), 61 – 83.
URL `http://pubsonline.informs.org/doi/abs/10.1287/msom.1060.0147`

Borders, M., Boyle, J., Turner, P., & Williams, M. (2009). Reducing ambulance diversion in California: Strategies and best practices. *California Healthcare Foundation*.

Boudreaux, E. D., & O'Hea, E. L. (2004). Patient satisfaction in the emergency department: A review of the literature and implications for practice. *The Journal of Emergency Medicine*, *26*(1), 13 – 26.
URL `http://www.sciencedirect.com/science/article/pii/S0736467903002853`

Bretthauer, K. M., Heese, H. S., Pun, H., & Coe, E. (2011). Blocking in healthcare operations: A new heuristic and an application. *Production and Operations Management*,

20(3), 375 – 391.
URL `http://dx.doi.org/10.1111/j.1937-5956.2011.01230.x`

Burström, L., Nordberg, M., Ornung, G., Castrén, M., Wiklund, T., Engstrom, M.-L., & Enlund, M. (2012). Physician-led team triage based on lean principles may be superior for efficiency and quality? A comparison of three emergency departments with different triage models. *Scandinavian Journal of Trauma Resuscitation & Emergency Medicine*, *20*, 57.

C, L., & Iyer, S. A. (2013). Application of queueing theory in health care: A literature review. *Operations Research for Health Care*, *2*(12), 25 – 39.
URL `http://www.sciencedirect.com/science/article/pii/S2211692313000039`

Campello, F., Ingolfsson, A., & Shumsky, R. A. (2013). Queueing models of case managers.

Carret, M. L. V., Fassa, A. C. G., & Domingues, M. R. (2009). Inappropriate use of emergency services: A systematic review of prevalence and associated factors. *Cadernos de Saúde Pública*, *25*(1), 7 – 28.

Ceschia, S., & Schaerf, A. (2011). Local search and lower bounds for the patient admission scheduling problem. *Computers & Operations Research*, *38*(10), 1452 – 1463.

Chalfin, D. B., Trzeciak, S., Likourezos, A., Baumann, B. M., Dellinger, R. P., & DELAY-ED study group (2007). Impact of delayed transfer of critically ill patients from the emergency department to the intensive care unit. *Critical Care Medicine*, *35*(6), 1477 – 1483.
URL `http://dx.doi.org/10.1097/01.CCM.0000266585.74905.5A`

Chan, C. W., Farias, V. F., Bambos, N., & Escobar, G. J. (2012a). Optimizing intensive care unit discharge decisions with patient readmissions. *Operations Research*, *60*(6), 1323 – 1341.
URL `http://pubsonline.informs.org/doi/abs/10.1287/opre.1120.1105`

Chan, C. W., Yom-Tov, G., & Escobar, G. (2012b). When to use speedup: An examination of service systems with returns. Tech. rep., Working Paper.

Chockalingam, A., Jayakumar, K., & Lawley, M. A. (2010). A stochastic control approach to avoiding emergency department overcrowding. In *Proceedings of the Winter Simulation Conference*, (pp. 2399 – 2411). IEEE.

Chonde, S. J., Ashour, O. M., Nembhard, D. A., & Kremer, G. E. O. (2013). Model comparison in emergency severity index level prediction. *Expert Systems with Applications*, *40*(17), 6901 – 6909.

Christ, M., Grossmann, F., Winter, D., Bingisser, R., & Platz, E. (2010). Modern triage in the emergency department. *Deutsches Ärzteblatt International*, *107*(50), 892.

Cochran, J., & Bharti, A. (2006). Stochastic bed balancing of an obstetrics hospital. *Health Care Management Science*, *9*(1), 31 – 45.
URL `http://dx.doi.org/10.1007/s10729-006-6278-6`

Cochran, J. K., & Roche, K. (2008). A queuing-based decision support methodology to estimate hospital inpatient bed demand. *The Journal of the Operational Research Society*, *59*(11), 1471 – 1482.
URL http://www.jstor.org/stable/20202231

Cohen, M., Hershey, J., & Weiss, E. (1980). Analysis of capacity decisions for progressive patient care hospital facilities. *Health Services Research*, *15*(2), 145 – 160.
URL http://europepmc.org/articles/PMC1072154

Cooney, D. R., Millin, M. G., Carter, A., Lawner, B. J., Nable, J. V., & Wallus, H. J. (2011). Ambulance diversion and emergency department offload delay: resource document for the national association of EMS physicians position statement. *Prehospital Emergency Care*, *15*(4), 555 – 561.

Cooney, D. R., Wojcik, S., Seth, N., Vasisko, C., & Stimson, K. (2013). Evaluation of ambulance offload delay at a university hospital emergency department. *International Journal of Emergency Medicine*, *6*(1), 1 – 4.

Crawford, E. A., Parikh, P. J., Kong, N., & Thakar, C. V. (2014). Analyzing discharge strategies during acute care: A discrete-event simulation study. *Medical Decision Making*, *34*(2), 231 – 241.
URL http://mdm.sagepub.com/content/34/2/231.abstract

Crawford, K., Morphet, J., Jones, T., Innes, K., Griffiths, D., & Williams, A. (2013). Initiatives to reduce overcrowding and access block in Australian emergency departments: A literature review.
URL http://www.sciencedirect.com/science/article/pii/S1322769613000930

Davies, R. (2007). "See and treat" or "see" and "treat" in an emergency department. In *Proceedings of the Winter Simulation Conference*, (pp. 1519–1522). IEEE.

de Bruin, A., Bekker, R., van Zanten, L., & Koole, G. (2010). Dimensioning hospital wards using the Erlang loss model. *Annals of Operations Research*, *178*(1), 23 – 43.
URL http://dx.doi.org/10.1007/s10479-009-0647-8

de Bruin, A., van Rossum, A., Visser, M., & Koole, G. (2007). Modeling the emergency cardiac in-patient flow: An application of queuing theory. *Health Care Management Science*, *10*(2), 125 – 137.
URL http://dx.doi.org/10.1007/s10729-007-9009-8

Defraeye, M., & Nieuwenhuyse, I. V. (2013). Staffing and scheduling under time-varying demand for service: A literature review. Tech. rep., KU Leuven, Research center for operations management.

Delgado, M. K., Meng, L. J., Mercer, M. P., Pines, J. M., Owens, D. K., & Zaric, G. S. (2013). Reducing ambulance diversion at hospital and regional levels: Systemic review of insights from simulation models. *Western Journal of Emergency Medicine*, *14*(5), 489.

Deo, S., & Gurvich, I. (2011). Centralized vs. decentralized ambulance diversion: A network perspective. *Management Science*, *57*(7), 1300 – 1319.

Derlet, R. W., & Richards, J. R. (2000). Overcrowding in the nation's emergency departments: Complex causes and disturbing effects. *Annals of Emergency Medicine*, *35*(1), 63 – 68.
URL `http://www.sciencedirect.com/science/article/pii/S0196064400701053`

Dobson, G., & Sainathan, A. (2011). On the impact of analyzing customer information and prioritizing in a service system. *Decision Support Systems*, *51*(4), 875 – 883.
URL `http://www.sciencedirect.com/science/article/pii/S0167923611000534`

Drupsteen, J., van der Vaart, T., & van Donk, D. P. (2013). Integrative practices in hospitals and their impact on patient flow. *International Journal of Operations & Production Management*, *33*(7), 912 – 933.

Durrande-Moreau, A. (1999). Waiting for service: Ten years of empirical research. *International Journal of Service Industry Management*, *10*(2), 171 – 194.

El-Darzi, E., Vasilakis, C., Chaussalet, T., & Millard, P. (1998). A simulation modelling approach to evaluating length of stay, occupancy, emptiness and bed blocking in a hospital geriatric department. *Health Care Management Science*, *1*(2), 143 – 149.
URL `http://ideas.repec.org/a/kap/hcarem/v1y1998i2p143-149.html`

Epstein, S. K., & Tian, L. (2006). Development of an emergency department work score to predict ambulance diversion. *Academic Emergency Medicine*, *13*(4), 421 – 426.
URL `http://dx.doi.org/10.1197/j.aem.2005.11.081`

Falvo, T., Grove, L., Stachura, R., Vega, D., Stike, R., Schlenker, M., & Zirkin, W. (2007). The opportunity loss of boarding admitted patients in the emergency department. *Academic Emergency Medicine*, *14*(4), 332 – 337.
URL `http://dx.doi.org/10.1197/j.aem.2006.11.011`

Fatovich, D. M., Nagree, Y., & Sprivulis, P. (2005). Access block causes emergency department overcrowding and ambulance diversion in Perth, Western Australia. *Emergency Medicine Journal*, *22*(5), 351 – 354.
URL `http://emj.bmj.com/content/22/5/351.abstract`

Fernandes, C. M., Price, A., & Christenson, J. M. (1997). Does reduced length of stay decrease the number of emergency department patients who leave without seeing a physician? *The Journal of Emergency Medicine*, *15*(3), 397 – 399.
URL `http://www.sciencedirect.com/science/article/pii/S0736467997000309`

Ferrin, D. M., Miller, M. J., & McBroom, D. L. (2007). Maximizing hospital financial impact and emergency department throughput with simulation. In *Proceedings of the 39th Conference on Winter Simulation: 40 Years! The Best is Yet to Come*, WSC '07, (pp. 1566 – 1573). Piscataway, NJ, USA: IEEE Press.
URL `http://dl.acm.org/citation.cfm?id=1351542.1351823`

FitzGerald, G., Jelinek, G. A., Scott, D., & Gerdtz, M. F. (2010). Emergency department triage revisited. *Emergency Medicine Journal*, *27*(2), 86 – 92.

Fomundam, S., & Herrmann, J. W. (2007). A survey of queuing theory applications in healthcare.

Forster, A. J., Stiell, I., Wells, G., Lee, A. J., & Van Walraven, C. (2003). The effect of hospital occupancy on emergency department length of stay and patient disposition. *Academic Emergency Medicine*, *10*(2), 127 – 133.
URL http://dx.doi.org/10.1197/aemj.10.2.127

Gallivan, S., & Utley, M. (2011). A technical note concerning emergency bed demand. *Health Care Management Science*, *14*(3), 250 – 252.

Garson, C., Hollander, J. E., Rhodes, K. V., Shofer, F. S., Baxt, W. G., & Pines, J. M. (2008). Emergency department patient preferences for boarding locations when hospitals are at full capacity. *Annals of Emergency Medicine*, *51*(1), 9 – 12.
URL http://www.sciencedirect.com/science/article/pii/S0196064407003770

Gilligan, P., Winder, S., Singh, I., Gupta, V., Kelly, P. O., & Hegarty, D. (2008). The boarders in the emergency department (BED) study. *Emergency Medicine Journal*, *25*(5), 265 – 269.
URL http://emj.bmj.com/content/25/5/265.abstract

Green, L. (2006). Queueing analysis in healthcare. In R. Hall (Ed.) *Patient Flow: Reducing Delay in Healthcare Delivery*, vol. 91 of *International Series in Operations Research & Management Science*, (pp. 281–307). Springer US.
URL http://dx.doi.org/10.1007/978-0-387-33636-7_10

Green, L. V. (2002). How many hospital beds? *Inquiry - Blue Cross and Blue Shield Association*, *39*, 400 – 412.
URL https://www.gsb.columbia.edu

Günal, M. M., & Pidd, M. (2010). Discrete event simulation for performance modelling in health care: A review of the literature. *Journal of Simulation*, *4*(1), 42 – 51.
URL http://www.ingentaconnect.com/content/pal/jos/2010/

Guo, P., & Zipkin, P. (2007). Analysis and comparison of queues with different levels of delay information. *Management Science*, *53*(6), 962 – 970.

Gurvich, I., & Van Mieghem, J. (2014). Collaboration in networks: Architectures, bottlenecks and capacity. Submitted to Manufacturing & Service Operation Management.

Hagtvedt, R., Ferguson, M., Griffin, P., Jones, G. T., & Keskinocak, P. (2009). Cooperative strategies to reduce ambulance diversion. In *Proceedings of the Winter Simulation Conference*, (pp. 1861–1874). IEEE.

Hall, R. W. (2006). *Patient Flow*. Springer.

Harper, P. R., & Shahani, A. K. (2002). Modelling for the planning and management of bed capacities in hospitals. *The Journal of the Operational Research Society*, *53*(1), 11 – 18.
URL http://www.jstor.org/stable/822874

Helm, J. E., AhmadBeygi, S., & Van Oyen, M. P. (2011). Design and analysis of hospital admission control for operational effectiveness. *Production and Operations Management*, *20*(3), 359 – 374.
URL http://dx.doi.org/10.1111/j.1937-5956.2011.01231.x

Helm, J. E., & Van Oyen, M. P. (2014). Design and optimization methods for elective hospital admissions. *Available at SSRN 2437936*.

Hemphill, R., & Nole, B. (2005). Relieving an overcrowded ED and increasing capacity for regional transfers: One hospital's bed management strategies. *Journal of Emergency Nursing: JEN : official publication of the Emergency Department Nurses Association*, *31*(3), 243 – 246.
URL http://dx.doi.org/10.1016/j.jen.2005.03.003

Higginson, I. (2012). Emergency department crowding. *Emergency Medicine Journal*, *29*(6), 437 – 443.
URL http://emj.bmj.com/content/29/6/437.abstract

Hobbs, D., Kunzman, S. C., Tandberg, D., & Sklar, D. (2000). Hospital factors associated with emergency center patients leaving without being seen. *The American Journal of Emergency Medicine*, *18*(7), 767 – 772.
URL http://www.sciencedirect.com/science/article/pii/S0735675700739301

Holm, L. B., & Barra, M. (2011). The consequences of how subject matter expert estimates are interpreted and modelled, demonstrated by an emergency department DES model comparing triangular and beta distributions. In *Proceedings of the Winter Simulation Conference*, WSC '11, (pp. 3654 – 3661). Winter Simulation Conference.
URL http://dl.acm.org/citation.cfm?id=2431518.2431952

Hoot, N. R., & Aronsky, D. (2008). Systematic review of emergency department crowding: Causes, effects, and solutions. *Annals of Emergency Medicine*, *52*(2), 126 – 136.
URL http://www.sciencedirect.com/science/article/pii/S0196064408006069

Hosseinifard, S. Z., Abbasi, B., & Minas, J. P. (2014). Intensive care unit discharge policies prior to treatment completion. *Operations Research for Health Care*, *3*(3), 168 – 175.
URL http://www.sciencedirect.com/science/article/pii/S221169231400040X

Howell, E., Bessman, E., Kravet, S., Kolodner, K., Marshall, R., & Wright, S. (2008). Active bed management by hospitalists and emergency department throughput. *Annals of Internal Medicine*, *149*(11), 804 – 810.
URL http://dx.doi.org/10.7326/0003-4819-149-11-200812020-00006

Howell, E., Bessman, E., Marshall, R., & Wright, S. (2010). Hospitalist bed management effecting throughput from the emergency department to the intensive care unit. *Journal of Critical Care*, *25*(2), 184 – 189.
URL http://www.sciencedirect.com/science/article/pii/S0883944109002287

Hwang, U., & Concato, J. (2004). Care in the emergency department: How crowded is overcrowded? *Academic Emergency Medicine*, *11*(10), 1097 – 1101.
URL http://dx.doi.org/10.1197/j.aem.2004.07.004

Hwang, U., McCarthy, M. L., Aronsky, D., Asplin, B., Crane, P. W., Craven, C. K., Epstein, S. K., Fee, C., Handel, D. A., Pines, J. M., Rathlev, N. K., Schafermeyer, R. W., Zwemer, F. L., Jr., & Bernstein, S. L. (2011). Measures of crowding in the emergency department: A systematic review. *Academic Emergency Medicine*, *18*(5), 527 – 538.
URL http://dx.doi.org/10.1111/j.1553-2712.2011.01054.x

Ibrahim, R., Armony, M., & Bassamboo, A. (2014). Does the past predict the future? The case of delay announcements in service systems. *submitted to Operations Research*. URL http://www.roubaibrahim.com/Delay%20Pred%20Reactions.pdf

Ibrahim, R., & Whitt, W. (2009a). Real-time delay estimation based on delay history. *Manufacturing & Service Operations Management*, *11*(3), 397 – 415. URL http://pubsonline.informs.org/doi/abs/10.1287/msom.1080.0223

Ibrahim, R., & Whitt, W. (2009b). Real-time delay estimation in overloaded multiserver queues with abandonments. *Management Science*, *55*(10), 1729 – 1742. URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1090.1041

Ibrahim, R., & Whitt, W. (2011a). Real-time delay estimation based on delay history in many-server service systems with time-varying arrivals. *Production and Operations Management*, *20*(5), 654 – 667. URL http://dx.doi.org/10.1111/j.1937-5956.2010.01196.x

Ibrahim, R., & Whitt, W. (2011b). Wait-time predictors for customer service systems with time-varying demand and capacity. *Operations Research*, *59*(5), 1106 – 1118.

Jayaprakash, N., O'Sullivan, R., Bey, T., Ahmed, S. S., & Lotfipour, S. (2009). Crowding and delivery of healthcare in emergency departments: The European perspective. *The Western Journal of Emergency Medicine*, *10*(4), 233 – 239. URL http://europepmc.org/articles/PMC2791723

Jensen, S., Lippert, A., & Østergaard, D. (2013). Handover of patients: A topical review of ambulance crew to emergency department handover. *Acta Anaesthesiologica Scandinavica*.

Jouini, O., Aguir, M. S., Akşin, Z., Karaesmen, F., & Dallery, Y. (2014). Data-based analysis of delay estimators and announcements in a call center.

Jouini, O., Akşin, Z., & Dallery, Y. (2011). Call centers with delay information: Models and insights. *Manufacturing & Service Operations Management*, *13*(4), 534 – 548. URL http://pubsonline.informs.org/doi/abs/10.1287/msom.1110.0339

Jouini, O., Dallery, Y., & Akşin, Z. (2009). Queueing models for full-flexible multi-class call centers with real-time anticipated delays. *International Journal of Production Economics*, *120*(2), 389 – 399. Special Issue on Introduction to Design and Analysis of Production Systems. URL http://www.sciencedirect.com/science/article/pii/S0925527309000140

Katsaliaki, K., & Mustafee, N. (2011). Applications of simulation within the healthcare context. *Journal of the Operational Research Society*, *62*(8), 1431 – 1451. URL http://www.ingentaconnect.com/content/pal/01605682/2011/00000062/00000008/art00001

Katz, K. L., Larson, B. M., & Larson, R. C. (1991). Prescription for the waiting-in-line blues: Entertain, enlighten, and engage. *Sloan Management Review*, *32*(2), 44 – 53.

Kc, D. S. (2014). Does multitasking improve performance? Evidence from the emergency department. *Manufacturing & Service Operations Management*, *16*(2), 168 – 183. URL http://pubsonline.informs.org/doi/abs/10.1287/msom.2013.0464

Kc, D. S., & Terwiesch, C. (2009). Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Science*, *55*(9), 1486 – 1498.
URL http://pubsonline.informs.org/doi/abs/10.1287/mnsc.1090.1037

Kc, D. S., & Terwiesch, C. (2012). An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management*, *14*(1), 50 – 65.
URL http://pubsonline.informs.org/doi/abs/10.1287/msom.1110.0341

Kelen, G. D., Kraus, C. K., McCarthy, M. L., Bass, E., Hsu, E. B., Li, G., Scheulen, J. J., Shahan, J. B., Brill, J. D., & Green, G. B. (2006). Inpatient disposition classification for the creation of hospital surge capacity: A multiphase study. *The Lancet*, *368*(9551), 1984 – 1990.
URL http://www.sciencedirect.com/science/article/pii/S0140673606698085

Kelen, G. D., Scheulen, J. J., & Hill, P. M. (2001). Effect of an emergency department (ED) managed acute care unit on ED overcrowding and emergency medical services diversion. *Academic Emergency Medicine*, *8*(11), 1095 – 1100.
URL http://dx.doi.org/10.1111/j.1553-2712.2001.tb01122.x

Kelton, W. D., Sadowski, R. P., & Sadowski, D. A. (2010). *Simulation with Arena*. New York, NY, USA: McGraw-Hill, Inc., 5 ed.

Kennedy, M., MacBean, C. E., Brand, C., Sundararajan, V., & McD Taylor, D. (2008). Review article: Leaving the emergency department without being seen. *Emergency Medicine Australasia*, *20*(4), 306 – 313.
URL http://dx.doi.org/10.1111/j.1742-6723.2008.01103.x

Khanna, S., Boyle, J., Good, N., & Lind, J. (2011). Impact of admission and discharge peak times on hospital overcrowding. *Studies in Health Technology and Informatics*, *168*, 82 – 88.
URL http://europepmc.org/abstract/MED/21893915

Khanna, S., Boyle, J., Good, N., & Lind, J. (2012). Unravelling relationships: Hospital occupancy levels, discharge timing and emergency department access block. *Emergency Medicine Australasia*, *24*(5), 510 – 517.
URL http://dx.doi.org/10.1111/j.1742-6723.2012.01587.x

Khare, R. K., Powell, E. S., Reinhardt, G., & Lucenti, M. (2009). Adding more beds to the emergency department or reducing admitted patient boarding times: Which has a more significant influence on emergency department congestion? *Annals of Emergency Medicine*, *53*(5), 575 – 585.
URL http://www.sciencedirect.com/science/article/pii/S0196064408014960

Koizumi, N., Kuno, E., & Smith, T. (2005). Modeling patient flows using a queuing network with blocking. *Health Care Management Science*, *8*(1), 49 – 60.
URL http://dx.doi.org/10.1007/s10729-005-5216-3

Kolb, E. M. W., Lee, T., & Peck, J. (2007). Effect of coupling between emergency department and inpatient unit on the overcrowding in emergency department. In *Proceedings of the 39th Conference on Winter Simulation: 40 Years! The Best is Yet to Come*,

WSC '07, (pp. 1586 – 1593). Piscataway, NJ, USA: IEEE Press.
URL http://dl.acm.org/citation.cfm?id=1351542.1351827

Kolb, E. M. W., Schoening, S., Peck, J., & Lee, T. (2008). Reducing emergency department overcrowding: Five patient buffer concepts in comparison. In *Proceedings of the 40th Conference on Winter Simulation*, WSC '08, (pp. 1516–1525). Winter Simulation Conference.
URL http://dl.acm.org/citation.cfm?id=1516744.1517010

Konrad, R., DeSotto, K., Grocela, A., McAuley, P., Wang, J., Lyons, J., & Bruin, M. (2013). Modeling the impact of changing patient flow processes in an emergency department: Insights from a computer simulation study. *Operations Research for Health Care*, *2*(4), 66 – 74.
URL http://www.sciencedirect.com/science/article/pii/S2211692313000052

Kuntz, L., Mennicken, R., & Scholtes, S. (2014). Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science*, *0*(0), null.
URL http://dx.doi.org/10.1287/mnsc.2014.1917

Lagoe, R. J., & Jastremski, M. S. (1990). Relieving overcrowded emergency departments through ambulance diversion. *Hospital Topics*, *68*(3), 23 – 27.

Larson, R. C. (1987). OR forum - Perspectives on queues: Social justice and the psychology of queueing. *Operations Research*, *35*(6), 895 – 905.
URL http://pubsonline.informs.org/doi/abs/10.1287/opre.35.6.895

Lim, M., Worster, A., Goeree, R., & Tarride, J.-E. (2013). Simulating an emergency department: The importance of modeling the interactions between physicians and delegates in a discrete event simulation. *BMC Medical Informatics and Decision Making*, *13*(1), 59.
URL http://www.biomedcentral.com/1472-6947/13/59

Lim, M. E., Nye, T., Bowen, J. M., Hurley, J., Goeree, R., & Tarride, J.-E. (2012). Mathematical modeling: The case of emergency department waiting times. *International Journal of Technology Assessment in Health Care*, *28*, 93 – 109.
URL http://journals.cambridge.org/article_S0266462312000013

Lin, D., Patrick, J., & Labeau, F. (2013). Estimating the waiting time of multi-priority emergency patients with downstream blocking. *Health Care Management Science*, (pp. 1 – 12).

Lin, W., Wu, Y., Zheng, J., & Chen, M. (2011). Analysis by data mining in the emergency medicine triage database at a Taiwanese regional hospital. *Expert Systems with Applications*, *38*(9), 11078 – 11084.

Lin, W.-T., Wang, S.-T., Chiang, T.-C., Shi, Y.-x., Chen, W.-y., & Chen, H.-m. (2010). Abnormal diagnosis of emergency department triage explored with data mining technology: An emergency department at a medical center in Taiwan taken as an example. *Expert Systems with Applications*, *37*(4), 2733 – 2741.

Litvak, N., van Rijsbergen, M., Boucherie, R. J., & van Houdenhoven, M. (2008). Managing the overflow of intensive care patients. *European Journal of Operational Research*, *185*(3), 998 – 1010.
URL http://www.sciencedirect.com/science/article/pii/S0377221706005819

Lovejoy, W. S., & Desmond, J. S. (2011). Littles law flow analysis of observation unit impact and sizing. *Academic Emergency Medicine*, *18*(2), 183 – 189.
URL http://dx.doi.org/10.1111/j.1553-2712.2010.00969.x

Luangkesorn, K. L., Bountourelis, T., Schaefer, A., Nabors, S., & Clermont, G. (2012). The case against utilization: Deceptive performance measures in inpatient care capacity models. In *Proceedings of the Winter Simulation Conference*, WSC '12, (pp. 76:1 – 76:12). Winter Simulation Conference.
URL http://dl.acm.org/citation.cfm?id=2429759.2429859

Luo, W., Cao, J., Gallagher, M., & Wiles, J. (2013). Estimating the intensity of ward admission and its effect on emergency department access block. *Statistics in Medicine*, *32*(15), 2681 – 2694.
URL http://dx.doi.org/10.1002/sim.5684

Mandelbaum, A., & Zeltyn, S. (2013). Data-stories about (im)patient customers in tele-queues. *Queueing Systems*, *75*(2-4), 115 – 146.
URL http://dx.doi.org/10.1007/s11134-013-9354-x

Marsh, A., Anderson, W., Bastani, A., & Bastani, A. (2004). Decreasing emergency department wait times for available inpatient beds by removing artificial variation. *Annals of Emergency Medicine*, *44*(4, Supplement), S28 – S29.
URL http://www.sciencedirect.com/science/article/pii/S0196064404008194

Marshall, A., Vasilakis, C., & El-Darzi, E. (2005). Length of stay-based patient flow models: Recent developments and future directions. *Health Care Management Science*, *8*(3), 213 – 220.

McManus, M. L., Long, M. C., Cooper, A., Mandell, J., Berwick, D. M., Pagano, M., & Litvak, E. (2003). Variability in surgical caseload and access to intensive care services. *Anesthesiology*, *98*(6), 1491 – 1496.
URL http://dx.doi.org/10.1097/00000542-200306000-00029

Medeiros, D. J., Swenson, E., & DeFlitch, C. (2008). Improving patient flow in a hospital emergency department. In *Proceedings of the Winter Simulation Conference*, WSC '08, (pp. 1526 – 1531). Winter Simulation Conference.
URL http://dl.acm.org/citation.cfm?id=1516744.1517011

Melton, N., Mitchell, M., Crilly, J., & Cooke, M. (2014). Patient characteristics and institutional factors associated with those who "did not wait" at a South East Queensland emergency department: Who are those who "did not wait" in ED? *Australasian Emergency Nursing Journal*, *17*(1), 11 – 18.
URL http://www.sciencedirect.com/science/article/pii/S1574626713001079

Michalowski, W., Kersten, M., Wilk, S., & Słowiński, R. (2007). Designing man–machine interactions for mobile clinical systems: MET triage support using palm handhelds. *European Journal of Operational Research*, *177*(3), 1409 – 1417.

Millard, W. B. (2011). Stand by to repel boarders: The rise of regional no-diversion policies. *Annals of Emergency Medicine*, *57*(5), A15 – A19.

Moloney, E. D., Bennett, K., ORiordan, D., & Silke, B. (2006). Emergency department census of patients awaiting admission following reorganisation of an admissions process. *Emergency Medicine Journal*, *23*(5), 363 – 367.
URL `http://emj.bmj.com/content/23/5/363.abstract`

Moskop, J. C., Sklar, D. P., Geiderman, J. M., Schears, R. M., & Bookman, K. J. (2009a). Emergency department crowding, part 1: Concept, causes, and moral consequences. *Annals of Emergency Medicine*, *53*(5), 605 – 611.
URL `http://www.sciencedirect.com/science/article/pii/S0196064408017836`

Moskop, J. C., Sklar, D. P., Geiderman, J. M., Schears, R. M., & Bookman, K. J. (2009b). Emergency department crowding, part 2: Barriers to reform and strategies to overcome them. *Annals of Emergency Medicine*, *53*(5), 612 – 617.
URL `http://www.sciencedirect.com/science/article/pii/S0196064408017885`

Mullen, P. M. (2003). Prioritising waiting lists: How and why? *European Journal of Operational Research*, *150*(1), 32 – 45.

Mustafee, N., Lyons, T., Rees, P., Davies, L., Ramsey, M., & Williams, M. D. (2012). Planning of bed capacities in specialized and integrated care units: Incorporating bed blockers in a simulation of surgical throughput. In *Proceedings of the Winter Simulation Conference*, WSC '12, (pp. 75:1 – 75:12). Winter Simulation Conference.
URL `http://dl.acm.org/citation.cfm?id=2429759.2429858`

Nie, W. (2000). Waiting: Integrating social and psychological perspectives in operations management. *Omega*, *28*(6), 611 – 629.
URL `http://www.sciencedirect.com/science/article/pii/S0305048300000190`

Olshaker, J. S., & Rathlev, N. K. (2006). Emergency department overcrowding and ambulance diversion: The impact and potential solutions of extended boarding of admitted patients in the emergency department. *The Journal of Emergency Medicine*, *30*(3), 351 – 356.
URL `http://www.sciencedirect.com/science/article/pii/S0736467905004427`

Oredsson, S., Jonsson, H., Rognes, J., Lind, L., Göransson, K. E., Ehrenberg, A., Asplund, K., Castrén, M., Farrohknia, N., et al. (2011). A systematic review of triage-related interventions to improve patient flow in emergency departments. *Scandinavian Journal of Trauma Resuscitation & Emergency Medicine*, *19*(43), 1 – 9.

Osorio, C., & Bierlaire, M. (2009). An analytic finite capacity queueing network model capturing the propagation of congestion and blocking. *European Journal of Operational Research*, *196*(3), 996 – 1007.
URL `http://www.sciencedirect.com/science/article/pii/S0377221708003871`

Ovens, H. (2011). ED overcrowding: The Ontario approach. *Academic Emergency Medicine*, *18*(12), 1242 – 1245.

Patel, P. B., Combs, M. A., & Vinson, D. R. (2014). Reduction of admit wait times: The effect of a leadership-based program. *Academic Emergency Medicine*, *21*(3), 266 – 273. URL `http://dx.doi.org/10.1111/acem.12327`

Patrick, J. (2011). Access to long-term care: The true cause of hospital congestion? *Production and Operations Management*, *20*(3), 347 – 358. URL `http://dx.doi.org/10.1111/j.1937-5956.2011.01229.x`

Paul, S. A., Reddy, M. C., & DeFlitch, C. J. (2010). A systematic review of simulation studies investigating emergency department overcrowding. *Simulation*, *86*(8-9), 559 – 571.

Peck, J. S., Benneyan, J. C., Nightingale, D. J., & Gaehde, S. A. (2012). Predicting emergency department inpatient admissions to improve same-day patient flow. *Academic Emergency Medicine*, *19*(9), 1045 – 1054. URL `http://dx.doi.org/10.1111/j.1553-2712.2012.01435.x`

Peck, J. S., Gaehde, S. A., Nightingale, D. J., Gelman, D. Y., Huckins, D. S., Lemons, M. F., Dickson, E. W., & Benneyan, J. C. (2013). Generalizability of a simple approach for predicting hospital admission from an emergency department. *Academic Emergency Medicine*, *20*(11), 1156 – 1163. URL `http://dx.doi.org/10.1111/acem.12244`

Pham, J. C., Patel, R., Millin, M. G., Kirsch, T. D., & Chanmugam, A. (2006). The effects of ambulance diversion: A comprehensive review. *Academic Emergency Medicine*, *13*(11), 1220 – 1227.

Pines, J. M., Batt, R. J., Hilton, J. A., & Terwiesch, C. (2011a). The financial consequences of lost demand and reducing boarding in hospital emergency departments. *Annals of Emergency Medicine*, *58*(4), 331 – 340. URL `http://www.sciencedirect.com/science/article/pii/S019606441100206X`

Pines, J. M., Hilton, J. A., Weber, E. J., Alkemade, A. J., Al Shabanah, H., Anderson, P. D., Bernhard, M., Bertini, A., Gries, A., Ferrandiz, S., Kumar, V. A., Harjola, V.-P., Hogan, B., Madsen, B., Mason, S., hln, G., Rainer, T., Rathlev, N., Revue, E., Richardson, D., Sattarian, M., & Schull, M. J. (2011b). International perspectives on emergency department crowding. *Academic Emergency Medicine*, *18*(12), 1358 – 1370. URL `http://dx.doi.org/10.1111/j.1553-2712.2011.01235.x`

Pines, J. M., Iyer, S., Disbot, M., Hollander, J. E., Shofer, F. S., & Datner, E. M. (2008). The effect of emergency department crowding on patient satisfaction for admitted patients. *Academic Emergency Medicine*, *15*(9), 825 – 831. URL `http://dx.doi.org/10.1111/j.1553-2712.2008.00200.x`

Plambeck, E. L., & Wang, Q. (2013). Implications of hyperbolic discounting for optimal pricing and scheduling of unpleasant services that generate future benefits. *Management Science*, *59*(8), 1927 – 1946. URL `http://dx.doi.org/10.1287/mnsc.1120.1673`

Powell, E. S., Khare, R. K., Venkatesh, A. K., Roo, B. D. V., Adams, J. G., & Reinhardt, G. (2012). The relationship between inpatient discharge timing and emergency department boarding. *The Journal of Emergency Medicine*, *42*(2), 186 – 196.
URL http://www.sciencedirect.com/science/article/pii/S0736467910006530

Rabin, E., Kocher, K., McClelland, M., Pines, J., Hwang, U., Rathlev, N., Asplin, B., Trueger, N. S., & Weber, E. (2012). Solutions to emergency department boarding and crowding are underused and may need to be legislated. *Health Affairs*, *31*(8), 1757 – 1766.
URL http://content.healthaffairs.org/content/31/8/1757.abstract

Ramirez, A., Fowler, J. W., & Wu, T. (2009). Analysis of ambulance diversion policies for a large-size hospital. In *Proceedings of the Winter Simulation Conference*, (pp. 1875 – 1886). Winter Simulation Conference.

Ramirez-Nafarrate, A., Fowler, J. W., & Wu, T. (2010). Bi-criteria analysis of ambulance diversion policies. In *Proceedings of the Winter Simulation Conference*, (pp. 2315 – 2326). IEEE.

Ramirez-Nafarrate, A., Fowler, J. W., & Wu, T. (2011). Design of centralized ambulance diversion policies using simulation-optimization. In *Proceedings of the Winter Simulation Conference*, (pp. 1251 – 1262). Winter Simulation Conference.

Ramirez-Nafarrate, A., Hafizoglu, A. B., Gel, E. S., & Fowler, J. W. (2014). Optimal control policies for ambulance diversion. *European Journal of Operational Research*, *236*(1), 298 – 312.
URL http://www.sciencedirect.com/science/article/pii/S0377221713009351

Richardson, D. B., & Mountain, D. (2009). Myths versus facts in emergency department overcrowding and hospital access block. *Medical Journal of Australia*, *190*(7), 369 – 374.

Robinson, D. J. (2013). An integrative review: Triage protocols and the effect on ED length of stay. *Journal of Emergency Nursing*, *39*(4), 398 – 408.
URL http://www.sciencedirect.com/science/article/pii/S0099176711006878

Saghafian, S., Austin, G., & Traub, S. (2014a). Operations research contributions to emergency department patient flow optimization: Review and research prospects. *Available at SSRN 2420163*.

Saghafian, S., Hopp, W. J., Van Oyen, M. P., Desmond, J. S., & Kronick, S. L. (2014b). Complexity-augmented triage: A tool for improving patient safety and operational efficiency. *Manufacturing & Service Operations Management*, *16*(3), 329 – 345.
URL http://dx.doi.org/10.1287/msom.2014.0487

Sharif, A. B., Stanford, D. A., Taylor, P., & Ziedins, I. (2014). A multi-class multi-server accumulating priority queue with application to health care. *Operations Research for Health Care*, *3*(2), 73 – 79. Special Issue of the 2012 conference of the EURO working group Operational Research Applied To Health Services (ORAHS).
URL http://www.sciencedirect.com/science/article/pii/S2211692314000034

Shi, P., Chou, M., Dai, J., Ding, D., & Sim, J. (2014). Models and insights for hospital inpatient operations: Time-dependent ED boarding time. *Management Science*.

Shone, R., Knight, V. A., & Williams, J. E. (2013). Comparisons between observable and unobservable M/M/1 queues with respect to optimal customer behavior. *European Journal of Operational Research*, *227*(1), 133 – 141.
URL `http://www.sciencedirect.com/science/article/pii/S0377221712009472`

Singer, A. J., Thode Jr, H. C., Viccellio, P., & Pines, J. M. (2011). The association between length of emergency department boarding and mortality. *Academic Emergency Medicine*, *18*(12), 1324 – 1329.

Sprivulis, P. C., Da Silva, J.-A., Jacobs, I. G., Frazer, A. R. L., & Jelinek, G. A. (2006). The association between hospital overcrowding and mortality among patients admitted via Western Australian emergency departments. *The Medical Journal of Australia*, *184*(5), 208 – 212.
URL `http://europepmc.org/abstract/MED/16515429`

Stanford, D., Taylor, P., & Ziedins, I. (2014). Waiting time distributions in the accumulating priority queue. *Queueing Systems*, *77*(3), 297 – 330.
URL `http://dx.doi.org/10.1007/s11134-013-9382-6`

Steele, R., & Kiss, A. (2008). EMDOC (emergency department overcrowding) internet-based safety net research. *The Journal of Emergency Medicine*, *35*(1), 101 – 107.
URL `http://www.sciencedirect.com/science/article/pii/S073646790700265X`

Sun, Y., Teow, K. L., Heng, B. H., Ooi, C. K., & Tay, S. Y. (2012). Real-time prediction of waiting time in the emergency department, using quantile regression. *Annals of Emergency Medicine*, *60*(3), 299 – 308.
URL `http://www.sciencedirect.com/science/article/pii/S0196064412002624`

Thompson, D. A., Yarnold, P. R., Williams, D. R., & Adams, S. L. (1996). Effects of actual waiting time, perceived waiting time, information delivery, and expressive quality on patient satisfaction in the emergency department. *Annals of Emergency Medicine*, *28*(6), 657 – 665.
URL `http://www.sciencedirect.com/science/article/pii/S0196064496700902`

Thompson, S., Nunez, M., Garfinkel, R., & Dean, M. D. (2009). OR practice - efficient short-term allocation and reallocation of patients to floors of a hospital during demand surges. *Operations Research*, *57*(2), 261 – 273.

Trzeciak, S., & Rivers, E. P. (2003). Emergency department overcrowding in the United States: An emerging threat to patient safety and public health. *Emergency Medicine Journal*, *20*(5), 402 – 405.
URL `http://emj.bmj.com/content/20/5/402.abstract`

van Ackere, A., Haxholdt, C., & Larsen, E. R. (2013). Dynamic capacity adjustments with reactive customers. *Omega*, *41*(4), 689 – 705.
URL `http://www.sciencedirect.com/science/article/pii/S0305048312001879`

Van der Vaart, T., Vastag, G., & Wijngaard, J. (2011). Facets of operational performance in an emergency room (ER). *International Journal of Production Economics*, *133*(1), 201 – 211.

van Essen, J. T., Houdenhoven, M., & Hurink, J. L. (2013). Clustering clinical departments for wards to achieve a prespecified blocking probability. Beta Working Paper WP-407, Beta Research School for Operations Management and Logistics, Eindhoven.

van Mieghem, J. A. (1995). Dynamic scheduling with convex delay costs: The generalized $c\mu$ rule. *The Annals of Applied Probability*, *5*(3), 809 – 833.
URL http://www.jstor.org/stable/2245125

Vance, J., & Sprivulis, P. (2005). Triage nurses validly and reliably estimate emergency department patient complexity. *Emergency Medicine Australasia*, *17*(4), 382 – 386.

Vanderby, S., & Carter, M. (2010). An evaluation of the applicability of system dynamics to patient flow modelling. *Journal of the Operational Research Society*, *61*(11), 1572 – 1581.

Vermeulen, M. J., Ray, J. G., Bell, C., Cayen, B., Stukel, T. A., & Schull, M. J. (2009). Disequilibrium between admitted and discharged hospitalized patients affects emergency department length of stay. *Annals of Emergency Medicine*, *54*(6), 794 – 804.
URL http://www.sciencedirect.com/science/article/pii/S019606440900465X

Vissers, J. M., Adan, I. J., & Dellaert, N. P. (2007). Developing a platform for comparison of hospital admission systems: An illustration. *European Journal of Operational Research*, *180*(3), 1290 – 1301.
URL http://www.sciencedirect.com/science/article/pii/S0377221706003377

Walsh, P., Cortez, V., & Bhakta, H. (2008). Patients would prefer ward to emergency department boarding while awaiting an inpatient bed. *The Journal of Emergency Medicine*, *34*(2), 221 – 226.
URL http://www.sciencedirect.com/science/article/pii/S0736467907003721

Welch, S., & Savitz, L. (2012). Exploring strategies to improve emergency department intake. *The Journal of Emergency Medicine*, *43*(1), 149 – 158.
URL http://www.sciencedirect.com/science/article/pii/S0736467911002824

Welch, S. J. (2009). Patient segmentation: Redesigning flow. *Emergency Medicine News*, *31*(8).

Welch, S. J. (2012). Using data to drive emergency department design: A metasynthesis. *Health Environments Research & Design Journal (HERD)*, *5*(3), 26 – 45.

Wiler, J. L., Bolandifar, E., Griffey, R. T., Poirier, R. F., & Olsen, T. (2013). An emergency department patient flow model based on queueing theory principles. *Academic Emergency Medicine*, *20*(9), 939 – 946.
URL http://dx.doi.org/10.1111/acem.12215

Wiler, J. L., Gentle, C., Halfpenny, J. M., Heins, A., Mehrotra, A., Mikhail, M. G., & Fite, D. (2010). Optimizing emergency department front-end operations. *Annals of Emergency Medicine*, *55*(2), 142 – 160.
URL http://www.sciencedirect.com/science/article/pii/S0196064409005319

Wong, H. J., Wu, R. C., Caesar, M., Abrams, H., & Morra, D. (2010). Smoothing inpatient discharges decreases emergency department congestion: A system dynamics simulation model. *Emergency Medicine Journal*, *27*(8), 593 – 598.
URL http://emj.bmj.com/content/27/8/593.abstract

Xu, K., & Chan, C. W. (2014). Using future information to reduce waiting times in the emergency department. Massachusetts Institute of Technology - Manuscript.

Yom-Tov, G. B., & Mandelbaum, A. (2014). Erlang-R: A time-varying queue with reentrant customers, in support of healthcare staffing. *Manufacturing & Service Operations Management*, *16*(2), 283 – 299.
URL http://dx.doi.org/10.1287/msom.2013.0474

Zeltyn, S., Marmor, Y. N., Mandelbaum, A., Carmeli, B., Greenshpan, O., Mesika, Y., Wasserkrug, S., Vortman, P., Shtub, A., Lauterman, T., Schwartz, D., Moskovitch, K., Tzafrir, S., & Basis, F. (2011). Simulation-based models of emergency departments: Operational, tactical, and strategic staffing. *ACM Transactions on Modeling and Computer Simulation*, *21*(4), 24:1 – 24:25.
URL http://doi.acm.org/10.1145/2000494.2000497

FACULTY OF ECONOMICS AND BUSINESS
Naamsestraat 69 bus 3500
3000 LEUVEN, BELGIË
tel. + 32 16 32 66 12
fax + 32 16 32 67 91
info@econ.kuleuven.be
www.econ.kuleuven.be