



HAL
open science

Simulation-based optimization of staffing levels in an emergency department

Karim Ghanes, Mathias Wargon, Oualid Jouini, Zied Jemai, Agapitos Diakogiannis, Romain Hellmann, Valérie Thomas, Ger Koole

► **To cite this version:**

Karim Ghanes, Mathias Wargon, Oualid Jouini, Zied Jemai, Agapitos Diakogiannis, et al.. Simulation-based optimization of staffing levels in an emergency department. *SIMULATION: Transactions of The Society for Modeling and Simulation International*, 2015, 91, pp.942-953. <10.1177/0037549715606808>. <hal-01265240>

HAL Id: hal-01265240

<https://hal.science/hal-01265240v1>

Submitted on 3 Feb 2016

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Simulation-Based Optimization of Staffing Levels in an Emergency Department

Karim Ghanes^a • Mathias Wargon^b • Oualid Jouini^a • Zied Jemai^c •

Agapitos Diakogiannis^a • Romain Hellmann^d • Valérie Thomas^d • Ger Koole^e

^a *Laboratoire Génie Industriel, Ecole Centrale Paris, Grande Voie des Vignes, 92290
Châtenay-Malabry, France*

^b *Emergency Department, Saint Camille Hospital, 2 Rue des Pères Camilliens, 94366
Bry-sur-Marne, France*

^c *OASIS -ENIT, University of Tunis El Manar, BP37 Le Belvedere, 1002 Tunis, Tunisia*

^d *Agence Régionale de Santé (ARS) Ile-de-France, 35 Rue de la Gare, 75019 Paris, France*

^e *Department of Mathematics, VU University Amsterdam, De Boelelaan 1081a, 1081 HV
Amsterdam, The Netherlands*

karim.ghanes@ecp.fr • M.Wargon@ch-bry.org • oualid.jouini@ecp.fr • zied.jemai@ecp.fr •
agapitos.dia@gmail.com • Romain.HELLMANN@ars.sante.fr • V.thomas@chmayotte.fr •
ger.koole@vu.nl

August 24, 2015

Abstract

In this case study, we use discrete event simulation to model and analyze a real-life emergency department (ED). Our approach relies on the appropriate integration of most real-life ED features to the simulation model in order to derive useful practical results. Data is supplied from the ED of the urban French hospital Saint Camille. Our purpose is to optimize the human resource staffing levels. We want to minimize the patient average length of stay (\overline{LOS}), by integrating the staffing budget constraint and a constraint securing that the most severe incidents will see a doctor within a specified time limit. The second constraint allows to avoid the perverse effect of only considering the \overline{LOS} metric that would delay the treatment of the most urgent patients. We use simulation-based optimization, in which we perform a sensitivity analysis expressing \overline{LOS} as a function of the staffing budget and also the average door-to-doctor time for urgent patients (\overline{DTDT}). We show that the budget has a diminishing marginal effect on the problem solution. Due to the correlation between \overline{LOS} and \overline{DTDT} , we also observe that the \overline{DTDT} constraint may significantly affect the feasibility of the problem or the value of the optimal solution.

Keywords: Emergency department, operations management, simulation, staffing optimization, length of stay, door-to-doctor time

1 Introduction

An emergency department (ED) is the main entrance to a hospital for emergency incidents, offering non-stop services for any kind of patients. The continuous increase in demand combined with austerity measures have led to extensive congestion [1]. There are several congestion signs, such as patients treated in corridors due to scarcity of accommodation resources, or patients experiencing excessive waiting times. Overcrowding has an immediate effect on the working environment of employees and on the quality of service offered to patients, which is mainly measured by certain key performance indicators (KPIs) [2].

Under a difficult economic context, ED managers are trying to improve performance by minimizing the mismatch between patient demand and supply. However, an ED is a complex environment with various types of heterogeneous patients and resources where most of the parameters are uncertain. Healthcare practitioners have therefore resorted to researchers in operations management and operations research in order to develop scientific approaches for the performance optimization of EDs. Their tools can be divided into two main categories: analytical methods and simulation. In this case study, the need for high impact solutions motivates us to use discrete event simulation (DES). This allows to capture most of the realistic features in an ED. In using simulation for ED operations management, we are following longstanding practice. [3], [4], [5], [6] and [7] conduct simulation studies for the analysis of EDs in Virginia (USA), London (Britain), Moncton (Canada), Kuwait and Dublin (Ireland) respectively. They address the problem of resource staffing optimization. [8] lay the foundation for developing a simulation tool to analyze the ED performance. For a background on simulation models for EDs, we refer the reader to the surveys by [9] and [10].

The simulation model proposed in this study is based on a comprehensive understanding of the real-world functioning of emergency departments. A field study was conducted for this purpose through a close collaboration with the ED of Saint Camille hospital. Saint Camille hospital is a teaching hospital situated in an Eastern suburb of Paris. Real data and expert judgments are both used for the construction of the model. For the validation, the model outputs are compared to historical data and judged by experts. In order to alleviate congestion, ED managers and the general management of Saint Camille hospital intend to invest in human staffing. Their objective is to improve the ED performance by investing in human resources. The question we are facing here is: By how much should the current staffing budget be increased and how should this additional budget be used in the allocation of human resources?

The selection of a KPI for ED optimization has always been a controversial subject. Neither the scientific community nor practitioners are able to decide about the most appropriate KPI, as each

indicator presents at the same time benefits and drawbacks. The most known and used KPI is the average length of stay (\overline{LOS}). LOS is the sum of the sojourn times in all subsections of the ED. It is the KPI on which EDs are generally judged in practice, because it allows to approach the ED in a holistic way. It is abundantly used in the literature as well. Some references include [11], [12], [13], [14], [15], [16], and [17]. However focusing only on \overline{LOS} could have important drawbacks. It gives an overview of the entire system performance but doesn't allow to figure out local strengths and weaknesses. Besides, the impact could be in the non-urgent cases, or worst, the non-urgent cases could be benefited on behalf of prolonging the waiting time of the urgent ones. From this appears the necessity to take another ED KPI into consideration, which is the Average door-to-doctor time (\overline{DTDT}). $DTDT$, also called time to first treatment or time to physician, describes the time between the patient's arrival and the first handling by a physician. $DTDT$ measures the most crucial element for seriously ill patients because they need urgent attention. For non urgent patients, the average $DTDT$ is generally close to the entire \overline{LOS} and thus the latter is sufficient as a KPI for this kind of patients. There are references in the literature that consider $DTDT$ as the sole performance indicator for the analysis of EDs. Examples include [18], [19] and [20]. Only rare papers such as [21] and [22] consider both indicators, as we do in this paper.

The main contributions of this paper can be summarized as follows. We propose a simulation model that is based on a comprehensive understanding of the ED functioning. Most common structural and functional characteristics of EDs, at least in France, are taken into consideration thanks to a close collaboration with Saint Camille ED. Based on the above, we point out a set of important ED features that are frequently ignored in the related literature. The model is close to the real system and is then appropriate to be used to address some operations management issues. We focus on the simulation-based optimization of staffing levels of the various human resource types involved in the ED. We study the effect of the staffing budget on \overline{LOS} , and show that it has a diminishing marginal effect. For instance, an increase of 10%, 20% and 30% in the staffing budget can generate an improvement of 33%, 44% and 50% in the optimal \overline{LOS} , respectively. We also show the effect of including a \overline{DTDT} constraint for urgent patients in the model. We investigate how this additional constraint affects the optimality and the feasibility of the staffing problem solution. The results point out the fact that considering \overline{DTDT} in addition to \overline{LOS} involves a trade-off that managers should be aware about. We also derive useful insights about which type of resource to prioritize according to the available budget and the \overline{DTDT} target. We surprisingly find that additional investments should be allocated in priority to doctors, which is counterintuitive to ED practitioners. Although the modeling is based on a specific ED, qualitative conclusions hold for other ED frameworks.

The rest of the paper is organized as follows. In Section 2.1, we describe how the ED characteristics are implemented in the simulation model and the way data is collected. In Section 2.2, we validate the simulation model using historical data and expert judgments. Furthermore, we highlight the detailed level of modeling and compare it with the existing literature. In Section 3, we conduct simulation-based optimization experiments for the ED staffing problem. In Section 4, we give concluding remarks and highlight some future research.

2 Emergency Department Modeling

In this section, we provide the building of the simulation model as well as its validation.

2.1 Simulation Model

We use Saint Camille hospital's ED as a main reference to build our model. In this section we give an overview of the service with its resources and processes as well as the necessary data to construct the simulation model.

Saint Camille hospital has approximately 300 beds and covers most of the medical and surgical specialties. Its ED is operating 24 hours per day and serves more than 60,000 patients per year. Within the ED, we consider the following different zones:

- The external waiting room for walk-in patient arrival
- The registration and triage zone
- A shock room (SR) for acute ill patients
- Examination rooms (ER) also called boxes or cubicles
- An internal waiting room with stretchers for lying patients
- An internal waiting room for sitting patients
- The Observation Unit (OU)

In addition, the ED includes an ambulance arrival area and a central operation room where all the tasks that do not require the presence of the patient are made, such as reporting on computer, interpretation of diagnostic tests, discussions between medical staff, preparation of equipments, etc.

Patients arriving to the ED cover a big range of severity levels. At the beginning of the process, patients are categorized by a triage nurse according to their condition into five degrees of severity, known as Emergency Severity Index (ESI), where ESI 1 are the most severe patients and ESI 5 the

least severe ones [23]. There are several different types of resources. The resources are also splited into dedicated groups for the ESIs, with different staffing levels for each group. A physician for instance can be either senior or junior. A junior physician can be responsible only for a combination of ESI 3, 4 and 5 patients, while seniors can treat all categories. There are also two different types of nurses: The first one, referred to as triage nurse, is dedicated to the triage. The other nurses are inside the ED and are in charge of in-process patients. Moreover, ESIs 1, 2 and 3 belong to a group of patients referred to as long circuit (LC) and are treated by dedicated physicians and nurses. ESIs 4 and 5 are part of a group called short circuit (SC) and are also treated by resources dedicated to them. The shock room is dedicated to ESI 1 patients and a part of ESIs 2 and 3 patients. The shock room is also known as trauma and resuscitation room [24, 14]. Examination rooms are also assigned to certain ESIs but with a different subdivision: medium boxes for ESIs 2 and 3, general boxes for ESI 4, and a fast track for ESI 5. Other resources such as stretcher bearers are not dedicated to any specific patient type. The reason for not including some resources in our model, such as janitorial staff, is that they do not really affect the system performance in terms of patient waiting times.

Similarly to [3], [13] and [5], our methodology is based on assessing the effect of staff changes on key performance indicators. We consider human and space resources in the model. Human resources are considered as control variables. The model development is performed using Arena simulation software provided by Rockwell Automation. During their sojourn, patients go over several stages that involve various types of limited resources, and then various patient waiting durations. The optimization of \overline{LOS} involves the optimization of the sum of these durations. Processing times such as physician examinations or diagnostic tests are considered as exogenous variables, and thus they are not to be optimized. The main waiting durations of the simulation model are given in Figure 1.

The patient path in the ED comprises a series of assessments that constitutes the ED process, as synthesized in Figure 1. Patients have different severity levels. Therefore, the process varies from one patient type to another. However, the typical complete patient stay in an ED can be divided into five main parts, as described below.

(1) From arrival to triage: Upon arrival to the ED, the patient is first registered at the reception desk and she is then triaged by the triage nurse in a dedicated box at the entry of the ED, based on the ESI triage system. The severity determines the priority of the patient over others [23] and how she will be routed to the appropriate resources throughout the process. When the triage nurse is busy, patients must wait in the external waiting room. The red code patients (ESI 1) generally

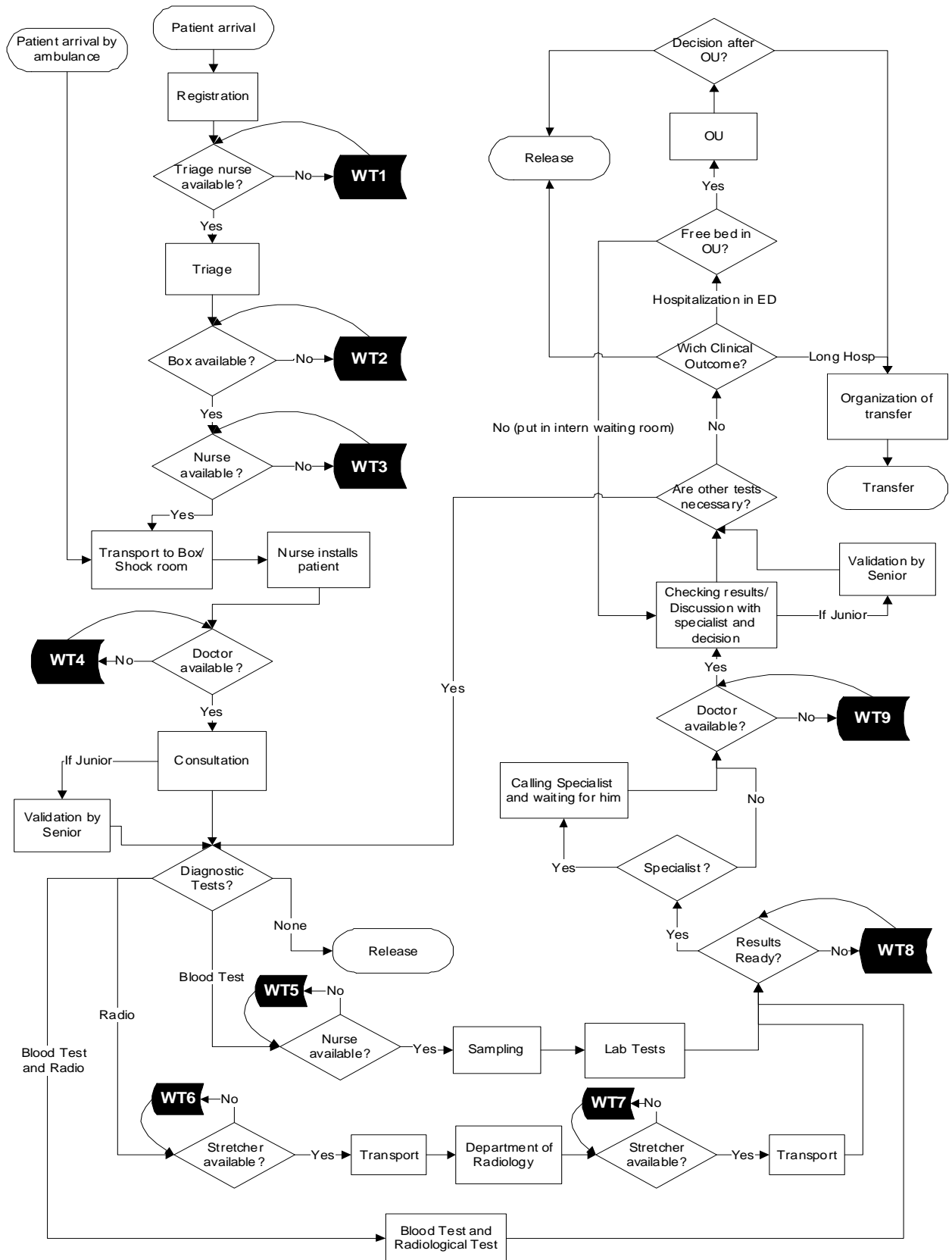


Figure 1: The conceptual model

arrive by ambulance. They must be stabilized immediately and skip triage.

(2) The initial consultation: After completing the triage process, the patient goes to the waiting room (sitting or on a stretcher depending on the severity) until an appropriate box becomes available. Then, she is transported and installed in the box by an appropriate nurse except ESI 5 patients who can do it themselves. The consultation starts once a doctor that is responsible for the patient's category becomes available. The doctor makes a first assessment and may request tests in order to confirm or refine her diagnosis. In case there is no examination required, the patient is discharged from the system. After the consultation, the doctor reports the diagnosis and the decisions made in the information system. Moreover, some important organizational aspects in the model are to be mentioned:

- Each decision made by a junior doctor must be validated by a senior one,
- Each patient must be treated by the same doctor and the same nurse all along the process. The "same patient-same staff" constraint, mentioned in [21] and [25], is a strong constraint with a significant impact on the system's behavior,
- Among any given ESI level and for any doctor, arriving patients have the priority over in-process ones.

(3) Diagnosis tests: According to the decision made by the doctor, there is a large variety of diagnosis tests that can follow the consultation. The doctor can order an electrocardiogram which is generally performed by a nurse in the box. Blood tests can be ordered; the nurse is responsible for the sampling in the box. Then, the sample is sent to the laboratory to be analyzed. During this time, the patient can wait in her box or can be put in an internal waiting room (if possible) in order to release the box and make it available for other patients. This decision depends on the patient's condition and we integrate it in our model by using a certain probability for each ESI. The duration of blood tests starts at that moment and finishes as soon as the results are ready. It represents one of the longest delays in the ED. Radiology tests can be also ordered with different combinations of X-Ray, CT-Scan, Echo and MRI. Note that LC patients must be transported by a Stretcher Bearer. When both tests are ordered, radiology and lab tests periods generally overlap. Analgesics can also be requested by the doctor. In the case of a perfusion, it will be done at the same time with the sampling (if any). It requires however an additional delay because a preparation beforehand is needed.

Diagnosis tests are undergone by resources located in another department and shared with other services of the hospital. Therefore, the durations that we fit do not represent only processing

times, but the total wait for the results. We include in this duration waiting times outside the ED. Consequently, reducing waiting times for external activities (radiology and laboratory) falls out of the scope of this study. They are considered as incompressible.

(4) Result interpretation and decision of the outcome: Once all the tests are completed, the doctor responsible for the patient evaluates the results, makes an interpretation and decides how the treatment procedure will be continued. In several cases, the doctor asks the patient to undertake supplementary examinations or even to redo some already taken examinations. The doctor can also request the opinion of a specialist from the hospital, a scenario that we model with a certain probability. Since the specialist belongs to another department, her intervention implies three additional durations: The time that the ED doctor spends to call the specialist by phone, the time necessary for the specialist to arrive, and the discussion with the ED doctor once she arrives. The duration is longer when the ED doctor is a junior one due to the lack of experience and her interest in learning.

(5) The process outcome: After the completion of the treatment procedure, the patient can be transferred to another service of the hospital, transferred to another hospital, admitted in the observation unit (OU) or discharged. When a patient is transferred to another department to be hospitalized, the responsible doctor must organize the transfer by phone. Then, the stretcher bearer is responsible for the transportation and the installation of the patient to the destination department. When a patient is transferred to another hospital, the responsible doctor must also call the hospital to organize the transfer. In this case, the transportation to the ambulance is done by the ambulance crew.

The OU is the area of the ED that hosts patients for a short stay before a transfer to another unit that could be the ward of the hospital or another hospital, or when the patient situation requires an additional observation before being released [26]. The beds are the critical resources of the OU. It has a limited capacity of beds and it admits and releases patients only during some specific periods of the day. Observation units are generally neglected in ED modeling in the literature, and yet it is very important to include them because they interact with the rest of the ED and have an impact on its performance. In Saint Camille ED, when the OU is full, patients supposed to be admitted are kept in the ED, laid in boxes or in the internal waiting room. In this case, a nurse from the ED must control these patients regularly, as described also in [27].

It is well known that the quality of output data relies on the accuracy of input parameters. Therefore, data collection and analysis are undertaken carefully. The first step consists of the

collection of the different types of data. In the second step, we model the data with statistical distributions in order to use them as input parameters for the model. Our simulation model requires three types of data: arrival pattern, processing times and routing probabilities. Depending on their type, ED data are more or less easy to collect. Thus we relied on the wide variety of data sources commonly used in similar studies and summarized in [9]: records from databases, interviews with experts and decision makers, and on-site observations; in addition to comparison with other hospitals' ED. Arrival pattern and some routing probabilities are relatively easy to collect since the corresponding data is systematically recorded and stored in the ED database. On the other hand, processing times and some process information are not recorded. For the above we used on-site observations and interviews with experts.

Arrival pattern: Similarly to [28] and [6], we assume that arrivals follow a non-homogenous Poisson process. The time dependent arrival pattern is quite typical for most EDs in the world [29]. Monday is usually the day that records the most arrivals, whereas higher arrival rates are found in the period between 10 am and 10 pm for any given day. Arrivals are modeled by using an average arrival rate $\hat{\lambda}(t)$ for each hour of the week (7 days \times 24 hours = 168 rates). These 168 rates are estimated from the database of Saint Camille's ED for 103 consecutive weeks, starting from September 2011 and ending in September 2013 (Figure 2).

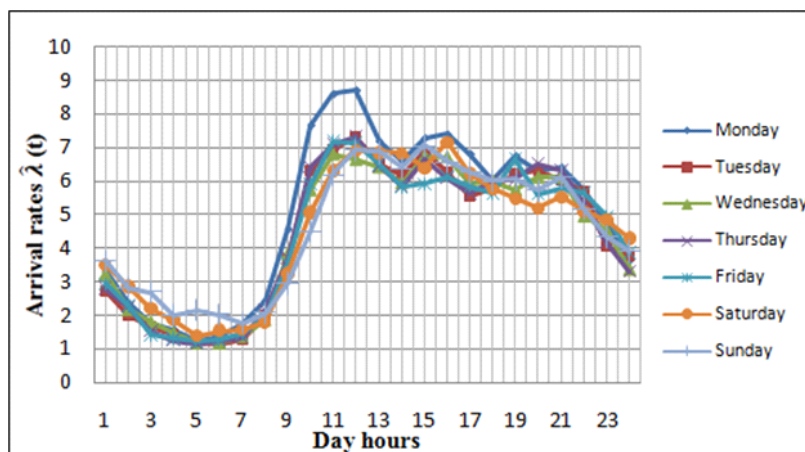


Figure 2: Estimated hourly patient arrival rates $\hat{\lambda}(t)$ per day

Processing times: There are 26 different service times that we modeled with statistical distribution fits, using the package Input Analyzer in Arena software. The processing times for each step of the process depend on the resource type (junior doctors are slower than seniors) as well as patient's category (critical patients require more time).

Routing probabilities: These probabilities depend on the patient’s ESI and represent the chance for a patient to experience or not a certain stage of the process. The probabilities needed in our model correspond for instance to diagnosis tests, the mix of these tests (imaging, lab test, none or both), imaging mix (X-ray, scan, echo or MRI), patient abandonment, the need for specialist opinion, the clinical outcome, Remaking tests, observation unit outcome, etc.

2.2 Model Validation

[30] explain that if the model is not a close approximation of the real system, any conclusions derived from the model are likely to be erroneous and may result in costly and ineffective decisions. Simulation models need to be built in a very precise way in order to represent the real environment as realistically as possible. The completion of our simulation model was a long procedure that contained many iterations; each step of the conceptual model had to be validated by experts in order to secure that it is an accurate representation of the system.

Exhaustivity: Concerning the granularity of simulation models, researchers have stated in the past that EDs are such complex systems that it is impossible to take all their features into consideration. [31] has shown that in most cases, 80% of model accuracy is obtained from only 20% of the model’s detail. However, ED models in the literature generally use many assumptions where important characteristics of the system are neglected. In most cases, such simplifications are more frequent in models using analytical methods, but they still exist in simulation models as well.

Building a realistic and useful simulation model requires an appropriate selection of the model’s level of detail. Table 1 synthesizes some of the important features included in our model, and compares that with the existing studies. For instance, the feature Resources Subdivisions refers to the differentiation of the staff members. As explained in [8], some EDs distinguish between acute and ambulatory patients and allocate doctors accordingly. Another possible subdivision is the difference between seniors and juniors (generally neglected). This is included in our model where processing times are function of both the expertise and the patient category.

Comparison with real data: To validate the simulation model, we compare between *LOS* given by our model and that obtained from the ED data using descriptive statistics.

We consider a steady-state type simulation run with one pseudo-infinite length of time during which the system is not re-initialized. This is coherent with the real system that works without interruption (24/7). The replication length is 11 weeks (110,880 minutes), of which one week is used as a warm-up period (10,080 minutes). The choice of the warm-up duration is based on graphical inspection of the time-series of the simulation outputs. We observe that after one week the system

Table 1: Comparison of previous works and the present study in terms of model’s granularity

	Centeno et al. (2003)	Komashie & Mousavi(2005)	Duguay & Chetouane (2007)	Ahmed & Alkhamis (2009)	Weng et al. (2009)	Present Study
Arrival process	Depends on day period	Depends on week day	Depends on week day	Depends on day hours	Depends on day period	Depends on week day and day hours
Patients’ categories	4	2	5	3	4	5
Included resources	Doctors Nurses Boxes	Doctors Nurses Boxes	Doctors Nurses Boxes	Receptionists Doctors Nurses Boxes Lab technicians Beds	Doctors Nurses Sick Beds	Stretcher Bearers Doctors Nurses Boxes Sick Beds Beds
Resources subdivision	No	Yes	No	No	Yes	Yes
Severity and/or expertise based processing times	Yes, based on severity	Yes, based on severity	Yes, based on severity	No	No	Yes, based on both
Lab tests/radiology	Yes	No	Yes	Yes	Yes	Yes
transportation times	No	No	No	No	No	Yes, for patients
Staff shifts	Yes	No	Yes	No	No	Yes
Teaching aspects	No	No	No	No	No	Yes
Specialist	No	No	No	No	No	Yes
Abandonment	Yes	No	No	No	No	Yes
Observation unit	No	Yes	No	Yes	Yes	Yes
Experiments	Simulation-optimization	Intuitive what-if scenarios	Intuitive what-if scenarios	Simulation-optimization	Simulation-optimization	Simulation-optimization
Control variables	Nurses	All included resources	All included resources	Doctors Nurses Lab technicians	Doctors Nurses	All included human resources

reaches typical conditions of steady-state situations. Note that we do not use a cool-down period because the ED works 24/7 without interruption.

Figure 3 provides a box-plot where the real *LOS* of 37,986 patients is compared to the *LOS* given by simulation for 7,604 patients. The outliers represent less than 5% for both real and simulated values. Figure 3 shows that there are some differences between the two distributions. Nevertheless, the comparison between the real and simulated cumulative distributions reveals encouraging similarities (Figure 4). For instance, starting from $LOS = 200$ minutes, the two distributions become very close. Furthermore, we successfully confronted two other indicators with expert judgment: resources workload and the durations of the five stages of the ED process (including the corresponding waiting durations). These encouraging similarities allowed considering the model reliable and valid to support experiments.

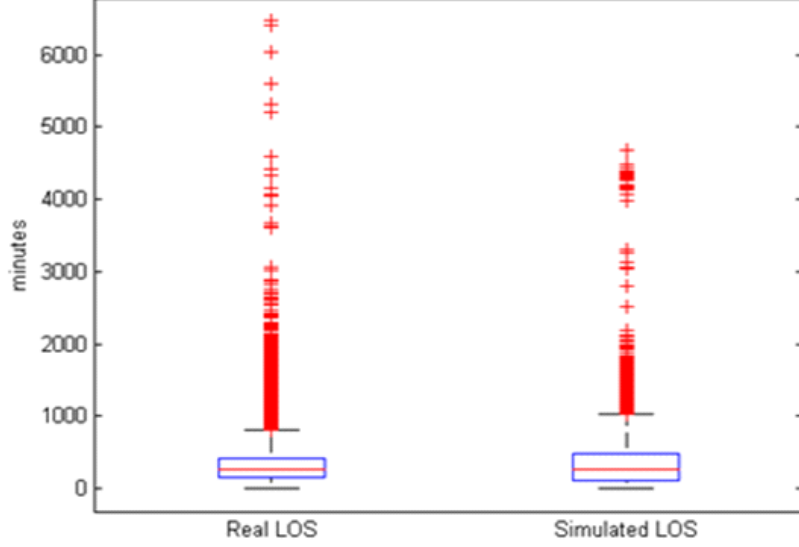


Figure 3: Real and simulated LOS

3 Staffing Level Optimization

Investing in human staffing is one of the possible ways to improve the ED performance. We want to address the following questions: By how much should we increase the current staffing budget, and how should this additional budget be used in the allocation of human resources? The results of this study has stood as a strong argument in order to convince the Saint Camille hospital management on the usefulness of increasing the funding for ED staffing. In general, similar approaches are also expected to support decision maker arbitrations.

We formulate an optimization problem that seeks to minimize the average length of stay under a budgetary constraint, and a constraint ensuring that the average $DTDT$ of LC patients (\overline{DTDT}) does not exceed some specified threshold. This is a hard problem, for which we use Arena OptQuest package for simulation optimization. OptQuest is a commercial global optimizer that uses heuristics to efficiently explore the set of feasible solutions [32, 33]. The ED uses two different shifts, a first one from 9:30 am to 6:30 pm (day shift), and another one from 6:30 pm to 9:30 am (night shift). Let $I = \{\text{Senior, Junior, Nurse, Triage nurse, Stretcher bearer}\}$ be the set of the considered resources with all possible subdivisions detailed in Section 2.1. Let $J = \{\text{Day shift, Night shift}\}$ be the set of the considered shifts. The real salaries of the ED staff have been used. The control variables $X_{i,j}$ represent the amount of a certain resource i during a given shift j . These variables are defined in Arena and used as control variables in OptQuest. For each resolution, OptQuest needs a starting solution that will serve as a starting point for exploring the set of feasible solutions. The initial parameters we choose correspond to the actual scheduling used in Saint Camille ED.

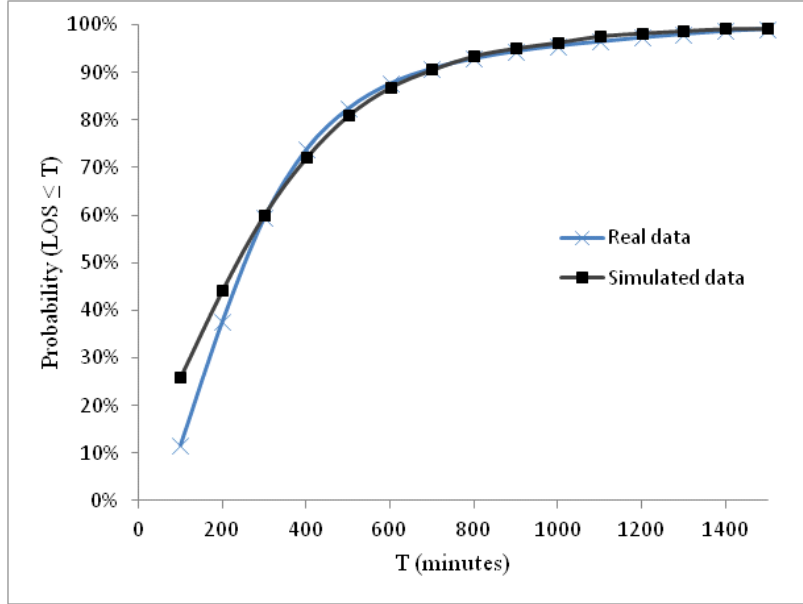


Figure 4: Cumulative distributions of real and simulated LOS

Since the results of the optimization can slightly vary according to the initial solution, we made each optimization several times by varying the starting parameter values. For practical reasons, the staffing levels for doctors during weekends will remain unchanged. The problem is expressed as follows:

$$\left\{ \begin{array}{l} \min \overline{LOS} \\ \text{subject to} \\ \sum_{i=1}^n \sum_{j=1}^m C_{i,j} X_{i,j} \leq C(1 + \alpha), \text{ for } i \in I, j \in J \\ \overline{DTDT} \leq L, \\ X_{i,j} \geq 0, \text{ for } i \in I, j \in J \end{array} \right. \quad (1)$$

where

\overline{LOS} = Average length of stay in the system,

$X_{i,j}$ = Amount of resource i during shift j ,

$C_{i,j}$ = Salary for resource i during shift j ,

C = Current staffing budget,

α = Percentage of additional staffing budget,

\overline{DTDT} = Average door-to-doctor time for LC patients (ESIs 1, 2 and 3),

$L = \overline{DTDT}$ limit.

The first constraint represents the staffing budget constraint. The budget limit is expressed as a function of a coefficient α that is the percentage of additional staffing budget. The second constraint

secures that the average door-to-doctor time for LC patients does not exceed a predetermined threshold L . Although the differences in staffing requirements for LC patients (junior doctors), we do only consider one single $DTDT$ constraint for all LC patient types. In practice, the most important point, with regard to $DTDT$, is the classification LC/SC and not the resource type allocations.

We perform a sensitivity analysis by varying at the same time α and L . Table 2 gives the results obtained by simulation optimization. Cells containing INF indicate that the combination of the budget and \overline{DTDT} constraints can not produce a feasible solution. The remaining values are the achieved \overline{LOS} , measured in minutes for an arbitrary patient. It should be mentioned that when the limit L is higher than 57 minutes, which is the value obtained in the initial simulation model with no supplementary budget, then the constraint is relaxed.

Table 2: Numerical experiments for the optimal \overline{LOS}

Additional Staffing Budget (α)	Current $\overline{DTDT}=57$	$\overline{DTDT}\leq 50$	$\overline{DTDT}\leq 40$	$\overline{DTDT}\leq 30$	$\overline{DTDT}\leq 20$	$\overline{DTDT}\leq 10$
0%	367	485	INF	INF	INF	INF
5%	323	389	397	INF	INF	INF
10%	246	277	277	INF	INF	INF
20%	205	205	205	229	INF	INF
30%	182	182	182	182	221	INF
40%	171	171	171	171	192	INF
50%	165	165	165	165	165	INF

Table 3: Resource staffing for optimal solutions of the sensitivity analysis

Additional Staffing budget (α)	Current $\overline{DTDT}=57$	$\overline{DTDT}\leq 50$	$\overline{DTDT}\leq 40$	$\overline{DTDT}\leq 30$	$\overline{DTDT}\leq 20$	$\overline{DTDT}\leq 10$
0%	Initial staffing	-1 senior SC night shift +1 junior ESI3 day shift +1 junior ESI45 day shift	INF	INF	INF	INF
	+1 senior SC day shift	+1 senior LC night shift -1 senior SC day shift +1 junior 345 night shift	+1 senior LC night shift -1 senior SC day shift +1 junior ESI45 day shift	INF	INF	INF
10%	+1 senior LC night shift	+1 senior LC night shift -1 senior SC night shift +1 junior 345 night shift +1 triage nurse day shift	+1 senior LC night shift -1 senior SC night shift +1 junior 345 night shift +1 triage nurse day shift	INF	INF	INF
	+1 senior LC night shift +2 nurses LC night shift	+2 senior LC night shift +1 junior ESI3	+2 senior LC night shift +1 junior ESI3	+1 senior LC night shift -1 senior SC day shift +1 junior ESI3	INF	INF
30%	+2 senior LC night shift	+2 senior LC night shift	+2 senior LC night shift	+2 senior LC night shift	+2 senior LC night shift	+2 senior LC night shift
	+1 senior SC day shift	+1 senior SC day shift	+1 senior SC day shift	+1 senior SC day shift	+1 senior SC day shift	+1 senior SC day shift
	+1 nurse LC night shift	+1 nurse LC night shift	+1 nurse LC night shift	+1 nurse LC night shift	+1 nurse LC night shift	+1 nurse LC night shift
	+1 triage nurse day shift	+1 triage nurse day shift	+1 triage nurse day shift	+1 triage nurse day shift	+1 triage nurse day shift	+1 triage nurses night shift

From Table 2, we observe that the budget has a diminishing marginal effect on performance. This can be seen from the first column of the table where the \overline{DTDT} constraint is relaxed. The highest marginal effect of the coefficient α on the \overline{LOS} corresponds to an investment of 10% of the current budget. This result allowed the ED managers with the general management of Saint Camille hospital to take an important tactical decision that consists on increasing the current staffing budget by 10% in order to reduce the current \overline{LOS} by 33%.

We also observe that the \overline{DTDT} constraint affects the optimality or the feasibility of the problem for small budgets. In certain cases, the limit L cannot be met by any possible allocation of resources and therefore the problem is infeasible. In other cases, by decreasing the limit of the \overline{DTDT} constraint for a certain budget, the optimal \overline{LOS} increases. For example, for $\alpha=20\%$, any value of $L \geq 40$ leads to an optimal \overline{LOS} of 205 minutes. However when $L = 30$, the optimal \overline{LOS} increases to 229 minutes. For high budget levels, the \overline{DTDT} constraint is automatically satisfied (staff allocation secures a low \overline{DTDT}), and thus the \overline{LOS} is independent of this constraint to some extent. This captures the trade-off between the two performance metrics.

The explanation of the last result requires the examination of the different solutions of Table 2 in terms of resource staffing. Table 3 provides the staffing changes for each optimal solution with regard to the initial staffing solution with no additional budget ($\alpha=0\%$, $L = 57$).

We can observe in all cases (for all problem formulations, i.e., with or without the \overline{DTDT} constraint) that the resource doctor is the most preferred one. There is always at least one additional doctor for all combinations of investment and \overline{DTDT} limit. Concerning the additional doctors' type, with the use of the \overline{DTDT} constraint ($L \leq 50$), resources tend to be devoted to LC patients in order to reduce \overline{DTDT} . For instance, when $\alpha=5\%$, an LC doctor is added during night shift to satisfy the \overline{DTDT} constraint while an SC doctor is added when this constraint is relaxed ($\overline{DTDT}=57$). This means that under the \overline{DTDT} constraint, there are less available resources for the SC patients (majority of patients) which increases the overall \overline{LOS} . Up to a certain budget ($\alpha=10\%$), there is no investment on other resources such as nurses. This is consistent with the fact that senior doctors' workload is the highest among all ED human resources.

When higher budgets are available, additional nurses are staffed. For instance, when $\alpha=20\%$, two additional nurses are added during night shift for LC patients when the \overline{DTDT} constraint is relaxed. Note that the nurse type privileged to overcome the \overline{DTDT} limit are triage nurses (not "in-process" nurses) because the triage stage and the corresponding waiting time is a part of the $DTDT$. For instance, when $\alpha=10\%$, one additional triage nurse is staffed during day shift to satisfy the \overline{DTDT} constraint. For higher budgets ($\alpha \geq 30\%$), resources are devoted independently of the \overline{DTDT} constraint. This means that regardless to the \overline{DTDT} constraint, there are enough resources

to secure that the LC patients will be treated within the threshold L .

The main conclusions from the above observations can be summarized as follows:

- Additional investments should be allocated in priority to doctors. A restrictive quality of service in terms of \overline{DTDT} will further give priority to LC doctors. This result seems surprising and counterintuitive to ED managers. As explained in [9], these findings are interesting given the large amount of research focusing on optimizing nursing allocation in various parts of the hospital [34, 35, 36]. Only rare papers focus on the important impact of doctor scheduling (compared to that of nurse) on the ED performance [37, 38].
- The lower is the budget, the more apparent is the correlation between \overline{LOS} and \overline{DTDT} .

4 Conclusions

We have built a realistic ED model using discrete event simulation. All common structural and functional characteristics of EDs, at least in France, were taken into consideration thanks to a close collaboration with practitioners. Based on the above, we point out a set of important ED features that are frequently ignored in the related literature. Although a simulation model can not be an exact imitation of the real system, the characteristics that we mention should be preferably taken into account in ED models, given their impact on the system performance. Our experiments focused on human staffing levels and provided useful insights to managers on the impact of the budget and \overline{DTDT} constraints on \overline{LOS} .

We observed that the staffing budget reveals a decreasing marginal effect on performance. For instance, an increase of 10%, 20% and 30% in the staffing budget can generate respectively an improvement of 33%, 44% and 50% in the optimal \overline{LOS} , when the \overline{DTDT} constraint is relaxed. Moreover, managers should be aware of the correlation between \overline{DTDT} and \overline{LOS} , for a given staffing budget. In some cases, \overline{DTDT} limits cannot be met with the use of several budgets, whereas in other cases meeting the \overline{DTDT} limits for the most severe patients has a negative effect on the total length of stay of all patients. The explanation lies in the fact that for low \overline{DTDT} targets, the budget tends to be devoted to urgent patients at the expense of non urgent patients (that represent the majority of patients) which affects the overall \overline{LOS} . Besides, we derived insights about the most appropriate type of resource to prioritize depending on the available staffing budget and the \overline{DTDT} target. We surprisingly find that additional investments should be allocated in priority to doctors, which is counterintuitive to ED practitioners. The results provide to managers a better understanding on how the budget can affect the system performance as well as on the

interdependency between the two main ED KPIs. This may then assist them in choosing the most appropriate operational decisions.

Some limitations of the current study are as follows. One limitation is related to input data. For instance, we considered routing probabilities and processing times as a function of the patient severity. However, in practice, some of these data depend also on the patient age or the medical specialty required for her treatment. Even though some correlations between several aspects exist, such as between ESI and age, we think that this represents a shortcoming. Moreover, we used an abandonment probability for patients as input to the model, while this parameter should be an output that depends on the patient time before abandonment. Unfortunately, the data about abandonment times is not reliable since it is not registered in the database when the patient leaves the ED, but only once her absence is noticed by the staff. Another limitation is related to the designed process. We assumed that the health status of a patient does not deteriorate during her sojourn in the ED, which is not the case in general. Since this may affect the in-process operations and durations, the simulation model can present a lack of accuracy.

References

- [1] N.R. Hoot and A. Aronsky. Systematic review of emergency department crowding: Causes, effects, and solutions. *Annals of Emergency Medicine*, 52(2):126–136, 2008.
- [2] K. Ghanes, A. Diakogiannis, O. Jouini, Z. Jemai, and M. Wargon. Key performance indicators for emergency departments: A survey from an operations management perspective. 2014. Working Paper, Ecole Centrale Paris.
- [3] M.D. Rossetti, G.F. Trzcinski, and S.A. Syverud. Emergency department simulation and determination of physician staffing schedules. *Proceedings of the 1999 Winter Simulation Conference*, 1999.
- [4] A. Komashie and A. Mousavi. Modeling emergency departments using discrete event simulation techniques. *Proceedings of the 37th Winter Simulation Conference*, pages 2681–2685, 2005.
- [5] C. Duguay and F. Chetouane. Modeling and improving emergency department systems using discrete event simulation. *Simulation*, 83(4):311–320, 2007.

- [6] M.A. Ahmed and T. M. Alkhamis. Simulation optimization for an emergency department healthcare unit in kuwait. *European Journal of Operational Research*, 198(3):936–942, 2009.
- [7] W. Abo-Hamad and A. Arisha. Simulation-based framework to improve patient experience in an emergency department. *European Journal of Operational Research*, 224:154–166, 2013.
- [8] D. Sinreich and Y. Marmor. Emergency department operations: The basis for developing a simulation tool. *IIE Transactions*, 37:233–245, 2005.
- [9] S.A. Paul, M.C. Reddy, and C.J. DeFlicht. A systematic review of simulation studies investigating emergency department overcrowding. *Simulation*, 86:559–571, 2010.
- [10] M.M. Günal and M. Pidd. Discrete event simulation for performance modelling in health care: a review of the literature. *Journal of Simulation*, 4(1):42–51, 2010.
- [11] J. Huang, B. Carmeli, and A. Mandelbaum. Control of patient flow in emergency departments, or multiclass queues with deadlines and feedback. 2012. Working Paper.
- [12] F. McGuire. Using simulation to reduce length of stay in emergency departments. *Journal of the Society for Health Systems*, 5(3):81–90, 1994.
- [13] M.A. Centeno, R. Giachetti, R. Linn, and A.M. Ismail. Emergency departments ii: a simulation-ilp based tool for scheduling er staff. *Proceedings of the 35th Conference on Winter Simulation: Driving Innovation*, pages 1930–1938, 2003.
- [14] S. Saghaian, W.J. Hopp, M.P. van Oyen, J.S. Desmond, and S.L. Kronick. Complexity-augmented triage: A tool for improving patient safety and operational efficiency. *Manufacturing and Service Operations Management*, 16(3):329–345, 2014.
- [15] M.H. Gorelick, K. Yen, and H.J. Yun. The effect of in-room registration on emergency department length of stay. *Emergency Medicine Journal*, 45(2):128–133, 2005.
- [16] J. Wang, J. Li, K. Tussey, and K. Ross. Reducing length of stay in emergency department: A simulation study at a community hospital. *IEEE Transactions on Systems, Man, and Cybernetics*, 42(6):1314–1322, 2012.
- [17] L.M. Song, A.L. Tucker, and K.L. Murell. The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. 2013. Working Paper.

- [18] M.W. Cooke, S. Wilson, and S. Pearson. The effect of a separate stream for minor injuries on accident and emergency department waiting times. *Journal of Emergency Medicine*, 19:28–30, 2012.
- [19] J.K. Cochran and K.T. Roche. A multiclass queuing network analysis methodology for improving hospital emergency department performance. *Computers Operations Research*, 36:1497–1512, 2009.
- [20] F.L. Lau and K.P. Leung. Waiting time in an urban accident and emergency department—a way to improve it. *Journal of Accident & Emergency Medicine*, 14(5):299–303, 1997.
- [21] S. Saghafian, W.J. Hopp, M.P. van Oyen, J.S. Desmond, and S.L. Kronick. Patient streaming as a mechanism for improving responsiveness in emergency departments. *Operations Research*, 60(5):1080–1097, 2012.
- [22] L. Burström, M. Nordberg, G. Örnung, M. Castrén, T. Wiklind, M. Engström, and M. Enlund. 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 and Emergency Medicine*, 20:1–10, 2012.
- [23] P. Tanabe, R. Myers, A. Zosel, J. Bricem, A.H. Ansari, J. Evans, Z. Martinovich, K.H. Todd, and J.A. Paice. Emergency department management of acute pain episodes in sickle cell disease. *Academic Emergency Medicine*, 14, 2007.
- [24] Y.H. Kuo, J.M.Y. Leung, and C.A. Graham. Simulation with data scarcity: Developing a simulation model of a hospital emergency department. *Proceedings of the 2012 Winter Simulation Conference*, pages 1–12, 2012.
- [25] C.E. Saunders, P.K. Makens, and L.J. Leblanc. Modeling emergency department operations using advanced computer simulation systems. *Annals of Emergency Medicine*, 18(2):134–140, 1989.
- [26] J.R. Broyles and J.K. Cochran. A queuing based statistical approximation of hospital emergency department boarding. *Proceedings of the International Conference on Computers and Industrial Engineering*, 2011.
- [27] S.J. Weng, B.C. Cheng, S.T. Kwong, L.M. Wang, and C.Y. Chang. Simulation optimization for emergency department resources allocation. *Proceedings of the 2011 Winter Simulation Conference*, pages 1231–1238, 2011.

- [28] G.B. Yom-Tov and A. Mandelbaum. Erlang-r: A time-varying queue with reentrant customers, in support of healthcare staffing. *Manufacturing & Service Operations Management*, 16(2):283–299, 2014.
- [29] G. Zayas-Caban, J. Xie, L.V. Green, and M. Lewis. Optimal control of an emergency room triage and treatment process. 2013. Working Paper.
- [30] A.M. Law and M.G. McComas. How to build valid and credible simulation models. *Proceedings of the 2001 Winter Simulation Conference*, pages 24–33, 2001.
- [31] S. Robinson. Simulation projects: building the right conceptual model. *Industrial Engineering*, 26(9):34–36, 1994.
- [32] B. Adenso-Diaz and M. Laguna. Fine-tuning of algorithms using fractional experimental designs and local search. *Operations Research*, 54(1):99–114, 2006.
- [33] J.P.C. Kleijnen and J. Wan. Optimization of simulated systems: Optquest and alternatives. *Simulation Modelling Practice and Theory*, 15(3):354–362, 2007.
- [34] H. E. Miller, W. P. Pierskalla, and G. J. Rath. Nurse scheduling using mathematical programming. *Operations Research*, 24(5):857–870, 1976.
- [35] L.J. Shuman, R.D. Jr. Spears, and J.P. Young. Operations research in health care: A critical analysis. *Baltimore, MD: Johns Hopkins University Press*, 1975.
- [36] E. K. Burke, P. De Causmaecker, G. vanden Berghe, and H. Van Landeghem. The state of the art of nurse rostering. *Journal of scheduling*, 7:441–499, 2004.
- [37] T.D. Jr. Clark and C.W. Waring. A simulation approach to analysis of emergency services and trauma center management. *Proceedings of the 19th Winter Simulation Conference*, pages 925–934, 1987.
- [38] G.W. Evans, T.B. Gor, and E. Unger. A simulation model for evaluating personnel schedules in a hospital emergency department. *Proceedings of the 28th Winter Simulation Conference*, pages 1205–1209, 1996.