



HAL
open science

A stochastic optimization model for shift scheduling in emergency departments

Omar El-Rifai, Thierry Garaix, Vincent Augusto, Xiaolan Xie

► **To cite this version:**

Omar El-Rifai, Thierry Garaix, Vincent Augusto, Xiaolan Xie. A stochastic optimization model for shift scheduling in emergency departments. *Health Care Management Science*, 2015, 18 (3), pp.289 - 302. 10.1007/s10729-014-9300-4 . hal-01905878

HAL Id: hal-01905878

<https://hal.science/hal-01905878v1>

Submitted on 26 Oct 2018

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.

A Stochastic Optimization Model for Shift Scheduling in Emergency Departments

Omar El-Rifai, Thierry Garaix, Vincent Augusto, Xiaolan Xie

*Center for Biomedical and Healthcare Engineering
Ecole Nationale Supérieure des Mines de Saint Etienne
Saint Etienne, France
email: xie@emse.fr*

Abstract

Excessive waiting time in Emergency Departments (ED) can be both a cause of frustration and more importantly, a health concern for patients. Waiting time arises when the demand for work goes beyond the facility's service capacity. ED service capacity mainly depends on human resources and on beds available for patients. In this paper, we focus on human resources organization in an ED and seek to best balance between service quality and working conditions. More specifically, we address the personnel scheduling problem in order to optimize the shift distribution among employees and minimize the total expected patients' waiting time. The problem is also characterized by a multi-stage re-entrant service process. With an appropriate approximation of patients' waiting times, we first propose a stochastic mixed-integer programming model that is solved by a sample average approximation (SAA) approach. The resulting personnel schedules are then evaluated using a discrete-event simulation model. Numerical experiments are then performed with data from a French hospital to compare different personnel scheduling strategies.

1 Introduction

Emergency departments are the crux of hospitals as they serve as a center for medical treatment and as a continuously operating portal for inpatient admissions [17]. It is important when considering the ED, to provide patients with timely health care while maintaining a healthy work environment for the personnel. Further ED are facing the challenge of providing a good quality of service while being cost-effective.

Increasing demand for health services in recent years and a reduction of the number of resources in the health industry due to budget cuts have led to chronic overcrowding in the ED [13]. From the patients' point of view, an overcrowded emergency department leads to excessive waiting times. While from a health specialist point of view, it can be the cause of an unfriendly work environment [16]. One of the major consequences of overcrowding with regards to the Operations Research community is a growing interest in developing strategies that make the most efficient use of the available resources [6]. In France, a recent report of the first national ED conference which gathered health practitioners and government officials decreed the urgency of the problem of long waiting times [1]. The report describes several strategies to reduce the potential bottlenecks that cause overcrowding which can also be found in the literature [20, 2, 3]. Among the recommendations, the authors mention the need to improve patient flow, the need to organize inpatients bed resources and the need to make a better use of available resources.

As ED work around the clock, a turnover of health care professionals is necessary in order to ensure an appropriate service level at all time. In the context of emergency departments, the work demand across the day is subject to the rate of arrivals of patients. In turn, the arrival rate of patients is subject to stochastic variations. These variations make the problem of organizing the workforce difficult.

A look at the number of patients and the staff level in Figure 1 observed at our collaborating hospital clearly shows mismatch of patients flow and personnel schedules. At every time period, the activity is estimated with the average number of patients present in the ED. Figure 1 clearly shows a peak of activity in the evenings. The distribution of staffs on duty does not appropriately match the peaks of activities observed in the ED. These observations prompted us to design a personnel scheduling model to meet the work demand accurately while satisfying the work constraints imposed.

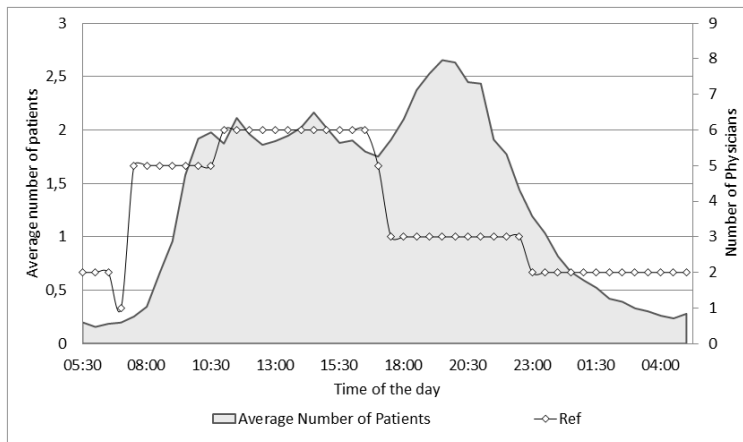


Figure 1: Preliminary analysis of the data from the pediatrics ED of Lille Hospital

In this paper, we present a human resources planning model for ED in order to explore the various human resources planning options and to take into account their robustness to face uncertainties and time-dependent patient arrival rate. The main efficiency criterion considered is the total patient waiting time. While other criteria are possible, total waiting time is a good indicator of patient service quality and ED crowding.

Personnel schedule is determined by a high-level stochastic programming model in which patient waiting times are approximated by a discrete time model and legal constraints for personnel scheduling are taken into account. The stochastic programming model is solved by a deterministic sample average approximation (SAA) [14] and mixed-integer programming. The resulting personnel schedules are evaluated with a more realistic ED discrete-event simulation model.

2 Workforce Scheduling in Emergency Departments

With the growing work constraints and the expansion of work facilities, the employee scheduling problem has become, for most industries, manually intractable. As early as 1954, mathematical models have been used in order to computationally assist industries to schedule their employees to work [5]. Employee scheduling is often decomposed into several sub-problems [6]. In a dynamic environment where the workload is variable, the first task is to be able to forecast the work demand for any scheduling horizon [15, 12, 13]. Once a work demand estimate is known, the decision maker has to determine an optimal staffing level for each scheduling period. After that, the staffing level is used as a basis for creating schedules that meet the staffing level appropriately. Finally, an assignment phase matches employees to the schedules created.

One of the peculiarities in scheduling studies conducted for health care systems is the focus on multi-skill scheduling. The work demand in health care systems needs to be addressed by appropriately qualified personnel and cannot be distributed arbitrarily between the resources [22, 10]. This multi-skill characteristic adds yet another layer of complexity to the workforce scheduling in ED that is often overlooked. We identify two main types of human resources in this study (physicians and nurses) and schedule them simultaneously. Simultaneous scheduling of different resources means having the work of one type of resource impact the schedules of the other.

With regards to the staffing problem, the work demand in the ED cannot be determined easily and depends on the arrivals of patients in the services and on the interactions between different resources. Literature on staff scheduling in the ED often considers stochastic demand patterns in order to address the uncertainty in the arrival of patients [9, 19]. Sinreich and Jabali [18] propose a simulation model coupled with an linear programming model to downsize emergency departments while keeping the patients' length of stay (LOS) constant. The simulation model dictates the required staffing profile while the linear programming model chooses the schedules that matches this staffing profile best. Zeltyn et al. [23] propose a unique simulation model to address the human resources allocation problem on multiple decision levels (operation, tactical and strategic). On the tactical level, they use two strategies: the Rough Cut Capacity Planing (RCCP) technique and the Offered Load approach. The later being a refinement of the former that allows to split the workload. The staffing level is computer hourly through the analysis of a simple queuing system. Queuing theory has also been extensively used for determining staffing levels because it allows the evaluation of quality of service criteria such as the waiting time of patients [7, 11]. However, analytic formulas for determining optimal staffing levels are often limited with regards to the complexity of the system they can analyze. As such, no analytical formula exists for optimally determining staffing level in systems with time-dependent stochastic arrival rates but instead numerical approximations have been used to alleviate this problem [21]. For example Ingolfsson et al. [8] use an integrated queuing and optimization model to investigate the effect of time-varying arrival rates in the context of workforce scheduling. The authors present a heuristic to construct potential schedule then use queuing theory to evaluate the resulting service levels. Izady and Worthington [9] use queuing theory and simulation in an iterative staffing scheme to determine schedules for an ED modeled as an infinite server network with different tasks and processes. After an appropriate staffing level is determined, an ILP is employed to find shifts that minimize the over-staffing and under-staffing. Cochran and Roche [4] suggest a queuing network approach for capacity planning of beds and staff in an ED. Their methodology is based on a multiple-customer class queuing analysis to increase ED access. Seasonal variation is of patient arrival patterns is considered.

In this work, the ED is represented as a multi-server multi-stage time-dependent priority queue system with re-entrance, Poisson distributed arrivals and exponential service times. Physicians and nurses are the servers whose cyclic schedules are to be determined. The planning horizon is discretized into time periods and a different mean arrival rate is given for every period of the day. We choose the expected total patient waiting time as the main criteria to optimize as it is often associated with patients' and staff well-being and is a good indicator of ED crowding. In order to address the complex nature of the problems, many studies addressed each sub-problem in the employee workforce scheduling independently by overlooking the inherent interaction between them [21]. As such, the optimal staffing level is often considered separately from the scheduling problem. This approach neglects complex waiting queues generated by uncertainties and changing patient arrival rate, which is crucial for ED crowding. In this work, we consider the staffing level as part of the scheduling problem and solve them both simultaneously. Integrating the two sub-problems allows us to schedule employees while reproducing the work dynamics of the system and simulating the patient flow.

The structure of the work is divided into two parts. First, we build an optimization model that finds staff schedules with operational constraints. Then, we evaluate more accurately our solution in a simulation model to measure the value of some of the assumptions in the optimization model. The optimization problem is solved by a Sample Average Approximation (SAA) solution methodology by considering different possible scenarios or realization of patient flows in the ED and optimizing the workforce scheduling problem considering all the different realizations. The remaining of the paper is organized as follows. We first describe the ED model in details in Section 3, then explain the optimization and the simulation models along with the assumptions that we consider in each in Sections 4 and 5 respectively. Then, we present numerical experiments in Section 6 and conclude in Section 7.

3 Emergency Department Queuing Network

In order to dynamically match the staff schedules and patient flow, we include in our system the dynamics of the patient flow through the ED. We present a queuing system with the key treatment steps and resources

of an ED. Patients can follow different trajectories through the ED depending on their pathology and their severity level. A simplified scheme for their trajectory is as follows. Patients are first assessed by a nurse to determine whether they need vital medical care, a step often called triage. If they need vital care, they are taken in priority to a resuscitation room. Otherwise, the patients wait in a waiting room until a physician is available to examine them. After the first assessment, patients are either directly treated by a nurse or are sent for auxiliary exams to confirm the diagnostic. The auxiliary exams range from a simple blood test to time-consuming imaging tests. After the auxiliary exams are performed, the patients are examined again by a physician and finally treated by a nurse before being discharged or admitted into another department in the hospital [9]. The second assessment leads to re-entrance of patients to the same resources. Priority is given to patients that are being assessed for the second time by physicians over patients arriving for the first time.

Each process in the ED can be modeled as a queue where jobs to be served are the patients and the hospital staff are the servers that handle these jobs. Figure 2 illustrates a basic queuing system. On the left side, we have patient arrivals into the system which follow an inhomogeneous Poisson distribution of parameter λ_t . Once in the system, they wait in a queue before being served by one of the servers of the system. The service time is assumed to be exponentially distributed in this paper. This corresponds to an $M_t/M/c$ queue model where M stands for the Markovian assumptions on the arrival and the service rates and c is the number of servers available.

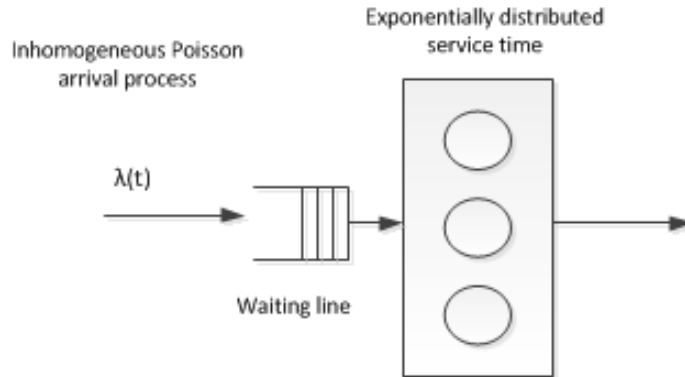


Figure 2: A basic queue model that represents a process in an ED

From the above, each ED has three main activities: assessment, treatment and auxiliary exams. The two health care providers that we model in the system are physicians and nurses. Physicians perform health assessments and examine the patients before and after having auxiliary exams. While nurses handle the final treatment administered to patients. As such, physicians are the servers for the two assessment queues of our system and nurses are the servers for the last queue. The overall system is illustrated in Figure 3. We do not schedule the lab technicians as they are often resources not entirely dedicated to the ED but instead delay patients that need to have auxiliary exams done.

Apart from the patients care pathway modeling, the queuing model also accounts for the stochastic nature of the work in the ED. The arrival rate of patients greatly varies during a day with peak arrival during the evenings and low arrival during the night. Service time needed to treat each patient can-not be known in advance. The stochastic service duration used in this paper models the different needs of patients. The number of patients arriving for each period of our planning horizon is parametrized by mean arrival rates that can be estimated from past data. Service distributions are estimated based on mean service times observed and found in the literature.

In Section 4, we present the mathematical formulation that we derived from the system specifications explained in this section. Some simplifying assumptions are made and detailed to solve the problem in reasonable time. Nonetheless, the solutions are evaluated using a more realistic simulation model whose specifications are explained in Section 5.

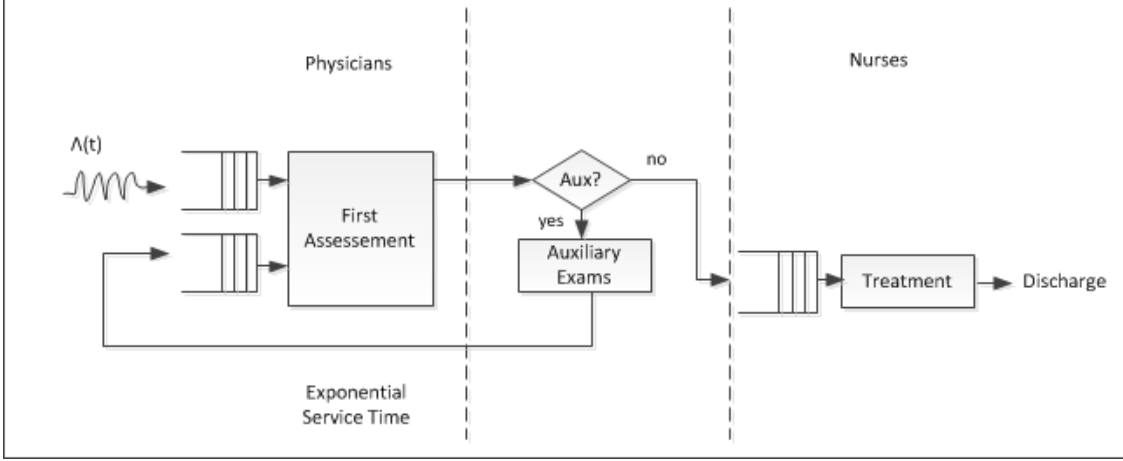


Figure 3: Process chart of patients in the ED

4 Mathematical Formulation

In this section, we present a mathematical formulation that is based on the details discussed in section 3. The objective is to optimize a waiting time performance indicator that we approximate with discrete time variables. As patients' waiting time is heavily influenced by the number of personnel present in the ED at any given time, the model finds optimal schedules over the planning horizon. For the time being, we are not interested in assigning the schedules to actual health personnel but only in finding an optimized schedule that meets the objective of minimizing the average total waiting time. In order to keep the model realistic, we allow for a limited number of different hours in which the staff can start working.

The number of patients arriving at each period t for each queue q is denoted $A_{t,q}(\omega)$. This symbol is overloaded in our model as it both designates the Poisson distributed arrival to the first two queues and the variable designating the arrival of patients to the nurses' queue. After the patients enter a queue, they are counted as waiting in the variable $W_{t,q}(\omega)$. Patients that are served in a queue, are counted in the variables $O_{t,q}(\omega)$. Then, parameter $L_{i,n,t}(\omega)$ determines the service capacity of the staff at each period. More specifically, the parameter determines the number of patients that n resources of type i can serve during period t . The ω index designates one scenario of patients' arrival and visit to the ED. The decision variables of the model $x_{i,n,t}$ define the number of health care providers working during period t .

To model the operational constraints, we use several parameters. A lower bound LBD and an upper bound UBD specify the possible lengths of the shifts that can be assigned to the personnel. The number of different hours of the day in which personnel of type i can start working is denoted by d_i . An upper bound TT_i is set on the number of total working hours that can be scheduled for resources of type i . The delay at the auxiliary exams is designated by δ . Finally some variables, $p_{i,t}$, $s_{i,t}$, $e_{i,t}$ and $z_{i,t}$ are used to facilitate the expression of shift constraints.

Some modeling assumptions are made in order to be able to solve the problem efficiently. A discrete-time approximation of the waiting time is calculated as $W_{t,q}$ minus $O_{t,q}$. What follows is that patients arriving during one period can only be treated in the next. Also, the waiting time of patients during the period they arrive is not counted. The patients are divided into two classes, those that do not require auxiliary exams and those that do. Each patients' class has a well defined trajectory from the start. This allows us to avoid having to route the patients to their appropriate queue after they have been assessed. This assumption is justifiable inasmuch as we do not care about each patients' trajectory individually. Instead, we keep track of the total number of patients in each queue of the system at each period t . We also assume the system is initially empty. We make this assumption by starting the scheduling horizon at an hour when it is reasonable to say that the system is clear. As of the data available to us, this is most true around 5:30 in the morning.

Furthermore, we do not distinguish between the different health care providers with the same skills and consider all physicians to be homogeneous and all nurses to be homogeneous. For the sake of simplicity, we omitted the possible visit of patients to the vital care unit. As of the data available from the UHC of Lille for the year 2011 and 2012, patients that require vital prognosis only constitute 1.79% of the total patients and those in need of resuscitation merely 0.45%. Finally, auxiliary exams are always assumed to last for a fixed duration. Of course, these assumptions are relaxed in the simulation model and evaluated for consistency in the numerical experiments.

| Parameter | Meaning | Domain |
|---------------------|---|---------------|
| t | The different periods of the day | $[1..T]$ |
| n | Number of resources of each type | $[1..N_i]$ |
| i | The different type of resources available | $\{1,2\}$ |
| q | The queue number | $[1..Q]$ |
| d_i | Maximum number of different starts for each resource | $[1..S_i]$ |
| TT_i | Number of working hours available to schedule for resource of type i | |
| LBD | Lower bound on the shift length allowable | $[1..47]$ |
| UBD | Upper bound on the shift length allowable | $[2..48]$ |
| δ | The delay at the auxiliary exams | |
| $A_{t,q}(\omega)$ | Poisson distributed arrivals in period t for queue q | $\mathbb{N}+$ |
| $L_{i,n,t}(\omega)$ | Sum of exponentially distributed consultation times for n resources of type i at period t | $\mathbb{N}+$ |

Table 1: The parameters of the scheduling model.

Objective:

$$\min_{\omega \in \Omega} \tilde{\mathbb{E}} \left[\sum_t \sum_q (W_{t,q}(\omega) - O_{t,q}(\omega)) \right] \quad (1)$$

Subject to:

$$W_{1,q}(\omega) = A_{1,q}(\omega) \quad \forall q \in \{1, 2\} \quad (2)$$

$$W_{1,q}(\omega) = 0 \quad \forall q \notin \{1, 2\} \quad (3)$$

| Variable | Meaning | Domain |
|-------------------|--|---------------|
| $x_{i,n,t}$ | = 1 if there are exactly n resources working at period t | $\{0,1\}$ |
| $p_{i,t}$ | Number of resources of type i present at period t | $[1..N_i]$ |
| $s_{i,t}$ | Number of resources of type i starting at period t | $[1..N_i]$ |
| $e_{i,t}$ | Number of resources of type i ending at period t | $[1..N_i]$ |
| $z_{i,t}$ | Indicator variable that is equal to 1 if at least one resource of type i starts working at t | $\{0,1\}$ |
| $W_{t,q}(\omega)$ | Number of patients that are waiting in queue q at the start of period t | $\mathbb{R}+$ |
| $O_{t,q}(\omega)$ | Number of patients in queue q that have been served during period t | $\mathbb{R}+$ |

Table 2: The variables of the scheduling model.

$$O_{1,q}(\omega) = 0 \quad \forall q \quad (4)$$

$$W_{t,1}(\omega) = W_{t-1,1}(\omega) + A_{t,1}(\omega) + O_{\max(t-\delta,1),2} - O_{t-1,1}(\omega) \quad \forall t \in [2..T] \quad (5)$$

$$W_{t,2}(\omega) = W_{t-1,2}(\omega) + A_{t,2}(\omega) - O_{t-1,2}(\omega) \quad \forall t \in [2..T] \quad (6)$$

$$W_{t,3}(\omega) = W_{t-1,3}(\omega) + O_{t-1,1}(\omega) - O_{t-1,3}(\omega) \quad \forall t \in [2..T] \quad (7)$$

$$O_{t,q}(\omega) \leq W_{t,q}(\omega) \quad \forall t \quad \forall q \quad (8)$$

$$\sum_{q \in \{1,2\}} O_{t,q}(\omega) \leq \sum_{n=1}^{N_1} L_{1,n,t}(\omega) * x_{1,n,t}(\omega) \quad \forall t \quad (9)$$

$$O_{t,3}(\omega) \leq \sum_{n=1}^{N_2} L_{2,n,t}(\omega) * x_{2,n,t}(\omega) \quad \forall t \quad (10)$$

$$\sum_t p_{i,t} \leq TT_i \quad \forall i \quad (11)$$

$$\sum_t (s_{i,t} - e_{i,t}) = 0 \quad \forall i \quad (12)$$

$$p_{i,t} = \sum_{t'=2}^t (s_{i,t'} - e_{i,t'}) + p_{i,1} \quad \forall t \in [1..T] \quad \forall i \quad (13)$$

$$p_{i,1} = p_{i,T} + s_{i,1} - e_{i,1} \quad \forall i \quad (14)$$

$$\sum_{t'=t+1}^{mod(t+LBD,T+1)} e_{i,t'} \leq p_{i,t} - s_{i,t} \quad \forall t \in [1..T] \quad \forall i \quad (15)$$

$$p_{i,t} \leq \sum_{t'=t+1}^{mod(t+UBD+1,T+1)} e_{i,t'} \quad \forall t \in [1..T] \quad \forall i \quad (16)$$

$$\sum_{n=1}^{N_i} n * x_{i,n,t} = p_{i,t} \quad \forall i \quad \forall t \quad (17)$$

$$\sum_{n=1}^{N_i} x_{i,n,t} = 1 \quad \forall i \quad \forall t \quad (18)$$

$$z_{i,t} * N_i \geq s_{i,t} \quad \forall i \quad \forall t \quad (19)$$

$$z_{i,t} \leq d_i \quad \forall i \quad \forall t \quad (20)$$

$$p_{i,t} \geq 1 \quad \forall i \quad \forall t \quad (21)$$

$$p_{i,t} \geq s_{i,t} + 1 \quad \forall i \quad \forall t \quad (22)$$

The objective function of the model (1) is a discrete time approximation of the total patients waiting time accumulated over all the scenarios. By summing over all the sample scenarios $\omega \in \Omega$ considered, we seek to obtain a good approximation of the estimated value of the actual waiting time of patients in the system. Constraints (3) and (4) initialize the number of patients in the queues in each scenario. This assumption of empty patient queue at the beginning of a day corresponds to our field observation of several ED in which often no patient remains late in the night such as at 5.30am. This is true whether the ED is busier during the day or not. The next set of equations of the problem are balance equations to track the movement of patients through the ED. As such, (5), (6) and (7) update the number of patients in each queue at every time period.

Constraints (8) specify a logical bound for the number of patients served during a period. This number can-not be larger than the number of patients already in the queue. Then, constraints (11) limit the total number of staffing hours available for each type of resource. Constraints (12) through (14) are logical constraints that define the range of values the variables p , s and e can take. Then, constraints (15) and (16) define the type of shifts that are allowed by specifying a lower bound and an upper bound on the shift lengths.

Constraints (17) and (18) assign a value to the variables x which is used to limit the number of patients served at every time period in constraints (9) and (10). Finally constraints (19) through (22) are operational constraints to make sure that there are at most d_i different moments of the day where resources of type i start working and that there is at least one resource of each type in every time period. We need to have an upper bound on the number of periods resources can start working in because physicians and nurses need to transmit patients' information to their colleagues before leaving the service. As such, constraints (22) make sure that the shifts of the different resources overlap.

5 Simulation Model

In order to evaluate the validity of the assumptions and simplifications made in the mathematical formulation, we test the solutions obtained using a simulation model. The simulation model includes several features not considered in the optimization model for a more accurate representation of the ED. The simulation improves over the mathematical formulation on several fronts: system specification, patients specification and resources activities.

As in the optimization model, exponential service time and Poisson arrival rates are used in the simulation model. However, the simulation uses a discrete event paradigm instead of the discrete time model paradigm in the optimization. This is important as one of the key assumptions in the optimization problem is that a discrete time model can serve to calculate patients' waiting time accurately. Second, the vital care pathway is no longer ignored. Patients requiring vital prognosis are attended by a physician and a nurse immediately. Auxiliary exams are divided into two categories: biological exams and diagnostic imaging exams. The biological exams make for 55% of the examinations.

Patients on the other hand are no longer homogeneous but are classified according to their acuity level (from 1 to 5 in increasing order of severity). The classification scheme used is that of the french hospitals known as CCMU (*Classification Clinique des Malades des Urgences*). This allows for a sickest-first prioritization in the different queues of the ED. Additionally, the patients that are classified as CCMU 4 or CCMU 5 skip the regular route and are treated in priority in the vital care pathway. These patients are served by both a nurse and a physician for a random duration.

Finally, with regards to resources, physicians are no longer solely occupied by patients. Instead, they also have to respond to administrative tasks such as searching patients' historical data or finding downstream beds for inpatients. Beds are not explicitly modeled in the system. instead, patients waiting in queues are considered to be assigned either a room or a hallway bed. Administrative tasks occur at random intervals but follow the same distribution laws as patient arrivals. Having the same distribution makes sense if we think that administrative tasks are more likely to occur when patients enter the system. Nurses no longer solely handle patients' treatments but also take care of the biological exams.

This improved ED model is used to evaluate the solutions obtained with the optimization model. Important system statistics such as the estimated patients' waiting time and the time spent in the system can be extracted and a confidence interval for these data can be calculated.

6 Numerical Experiments

Numerical experiments have been conducted to show the potential reduction of patients' waiting time in an ED by adjusting the workforces' shift schedules. The parameters settings for the optimization model, the simulation model, and the real-world benchmark are described in Section 6.1. In Section 6.2, we check that the optimization model matches enough key features of the queuing system to provide suitable solutions. In order to do so, the schedules obtained by running the optimization model are evaluated and compared to simulation tests. The performances of different scheduling strategies are also compared. The strategies differ in the constraints on the employees' shifts. In Section 6.3 the robustness of schedules facing significant demand increase is evaluated and in Section 6.4 the transition from week schedules to week-end schedules is investigated. The optimization model is solved using a commercial solver (IBM Ilog Cplex 12.5). All tests were performed on an Intel(R) Xeon(R) 4 cores CPU E5520 @ 2.27GHz with 8 MB of cache memory and 8 GB of RAM. The evaluation is done using AnyLogic 6.9.

6.1 Parameters Settings and Real-World Data

The parameters used in the experimental section are based on data for the years 2011-2012 of patients' entries in the pediatrics' ED of the University Hospital Center (UHC) of Lille France. We extracted from these data an average number of patients' arrivals per half-hour of the day and used them as parameters for the Poisson distributions of each period. The length of the period also called time step $\Delta = 30$ minutes is chosen

primarily to coincide with the shifts change marks used in the hospital. Also, the patients’ arrival averages are calculated according to length of the time step. Average values would be less significant if they relied on less data. As the data available only contain arrival and departure times, we did not determine the service duration of patients in the same manner as the arrivals. Instead, we relied on average service durations found in the literature and on experts’ opinion to calibrate these parameters. However, the patients’ lengths of stays obtained with these service durations are of the same order of magnitude than those observed in the real data. Based on these parameters, we generate several possible scenarios for patients arrivals and service time.

In the UHC of Lille, the number of resources working during the week differs drastically from the number of resources working on the weekends. Therefore, two set of optimization tests are run. The first set of tests uses the total staff hours available during the week and the arrival distributions extracted for weekdays and the second set of tests relies on data for weekends. The total staff hours available serves as an upper bound on the number of hours scheduled. Each scenario in the optimization model is generated by obtaining a realization ω of the distributions of the arrival and service times. Once the different parameters of the models are set, the scenarios are the only input to the model.

After the schedules are optimized, we evaluate these schedules and the waiting time they produce in the simulation model. As the schedules we obtain are cyclical, the evaluation is done for several consecutive days. We introduce a warm-up period in order to reach stable results. Evaluating the performance of the schedules can be done on a large number of replications which allows the calculation a confidence interval. The parameters used for the experimental runs are summarized in Table 3.

To validate the pertinence of a scenario-based strategy, we ran three different instances of optimization tests. After obtaining the resulting schedules from each test, we evaluated the evolution of the expected waiting time in the simulation model using the resulting schedules. We see the results in Figure 4, where the average waiting time over all three tests are displayed. The average waiting time of the patients decreases with the increase of the number of scenarios used in the optimization. We notice that the expected waiting time stabilizes for about $|\Omega| = 100$ scenarios.

In terms of computational feasibility, our model strongly depends on the number of scenarios used for optimization and on the time step chosen. This is because the number of variables increases linearly with the number of scenarios and the number of periods. Fortunately, we showed that the simulated average waiting time converges with solutions obtained from 100 scenarios. From our tests, all solutions with 100 scenarios were obtained within 4 hours with a 0.11% relative gap in the worst case.

Other variables in the model also influence the running time but to a lesser extent. As such, the number of patients in the queues does not increase the number of variables but only the range the variables can take. On the other hand, the number of variables representing resources: $x_{i,n,t}$ increases as the number of resources increases but is independent of the number of scenarios. For example, in a large ED with about 200 patients daily and 10 to 15 resources of each type, the optimal solution is still obtained within the 4 hours of computation time.

| Parameter | Definition | Value |
|-----------------|--|------------------|
| L | Length of scheduling horizon | 24 hours |
| Δ | Length of period in the scheduling horizon | 30 mins |
| Λ | Mean number of patients entering the system in one day | 65(week), 66(we) |
| $\frac{1}{\mu}$ | Mean service time per period | 30 mins |
| σ | Number of scenarios used in the optimization | 100 |
| τ | The time limit set for the optimization model | 4 hours |

Table 3: Parameters of the system and of the optimization algorithm

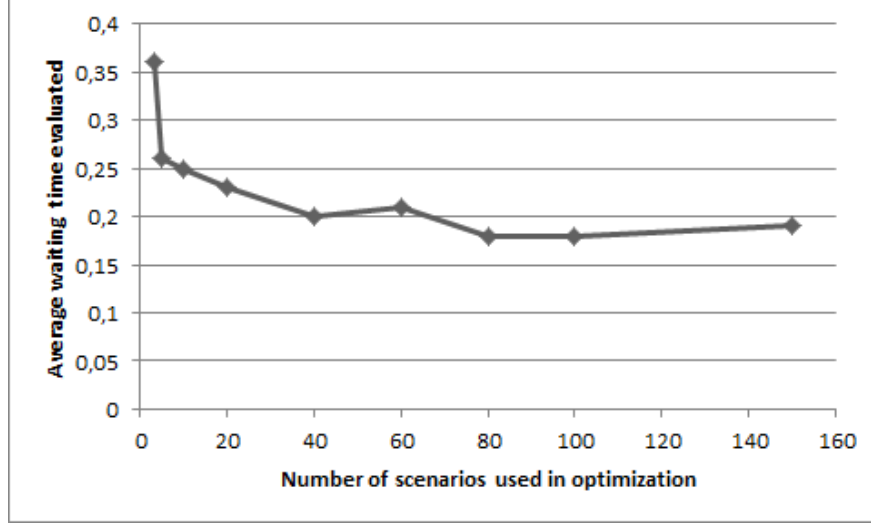


Figure 4: The evolution of waiting time with the number of scenarios used in the optimization model

Throughout the tests, the schedules obtained are compared against the reference schedule *ref* that is currently put in place in the ED of Lille. The first alternative we test are schedules obtained with shifts whose length is fixed to 8 hours for both physicians and nurses. The schedules obtained with this strategy give us an idea of how much we can improve by changing the starting times along with more restrictive shift constraints than what is currently practiced. Then, we keep the shifts lengths range from 4 to 12 hours as it is the case in the UHC of Lille but modify the starting times. The third and final alternative is to create shifts with no constraint on their lengths. In this last case, the only constraint is the total number of hours worked. This last strategy is not realistic but merely allows us to compute a lower bound on the total waiting time of the patients in the optimization model. Constraints on every schedule strategies are given in Table 4. The parameters d_i and TT_i refer to the parameters presented in the optimization model.

| Shifts | $LBD(h)$ | $UBD(h)$ | d_1 | d_2 | TT_1 | TT_2 |
|------------|----------|----------|----------------------|-------|---------------------------|--------|
| <i>ref</i> | 4 | 12 | 7 (week), 5(weekend) | 6 | 178 (week), 130 (weekend) | 124 |
| (4-12h) | 4 | 12 | 7 (week), 5(weekend) | 6 | 178 (week), 130 (weekend) | 124 |
| (8h) | 8 | 8 | 7 (week), 5(weekend) | 6 | 178 (week), 130 (weekend) | 124 |
| (1 – 24h) | 1 | 24 | 7 (week), 5(weekend) | 6 | 178 (week), 130 (weekend) | 124 |

Table 4: Parameters for the scheduling strategies

6.2 Analysis of optimized schedules

We verify in a first part that the optimized schedules obtained can actually reduce the average waiting time of patients compared to the reference schedule when tested in simulation conditions. In order to do so, we allow for different assumptions on the shifts as explained previously. In all the alternatives tested, we allow the shifts to start on any period of the day but only permit a limited number of shift changes to occur as it is currently the case. Our analysis in this subsection focuses on weekends but results obtained for weekdays are similar and lead to the same conclusions.

Table 5 summarizes the results obtained for schedules optimized to weekend conditions. The first column specifies the shifts type used. The number next to the shift type specifies different runs of the optimization problem, each with a different set of scenarios. $W - O$, is total objective value converted to average hour units per patient. This value gives an approximation on the expected waiting time as computed in the optimization model. The total waiting time associated with each queue, q in the optimization model is given

| Shifts | W-O (h) | $W_1 - O_1$ (h) | $W_2 - O_2$ (h) | $W_3 - O_3$ (h) | Gap % | Time(mins) |
|-------------|---------|-----------------|-----------------|-----------------|-------|------------|
| ref (1) | 0.41 | 1730 | 2581 | 1131 | 0.00 | 0 |
| (2) | 0.39 | 1767 | 2175 | 1200 | 0.00 | 0 |
| (3) | 0.47 | 1919 | 2939 | 1289 | 0.00 | 0 |
| 8hrs (1) | 0.27 | 1449 | 1415 | 683 | 0.02 | 240 |
| (2) | 0.26 | 1416 | 1311 | 738 | 0.01 | 240 |
| (3) | 0.31 | 1645 | 1690 | 735 | 0.11 | 240 |
| 4-12hrs (1) | 0.21 | 1283 | 1101 | 431 | 0.00 | 20 |
| (2) | 0.20 | 1143 | 1133 | 415 | 0.00 | 34 |
| (3) | 0.24 | 1334 | 1356 | 522 | 0.00 | 42 |
| 1-24hrs (1) | 0.21 | 1236 | 1109 | 469 | 0.00 | 180 |
| (2) | 0.20 | 1112 | 1130 | 400 | 0.00 | 53 |
| (3) | 0.24 | 1340 | 1342 | 523 | 0.00 | 51 |

Table 5: Optimization results for the weekend scenarios

in the columns $W_q - O_q$, for $q \in \{1, 2, 3\}$. Next, the Gap indicates the percentage distance between best lower and upper bounds obtained. An time limit of 240 minutes is used for each run of the optimization problem.

The simulation model of Section 5, is used to evaluate each solution schedule. The results of the evaluation are presented in Table 6. The simulation is run over 100 replications each of 10 successive days. Since we only have a unique reference schedule, only one row keeps track of the expected waiting time using the reference schedule. Table 6 first notes the patients' expected waiting time in *Wait*. The next columns indicated the expected and max waiting time in the physicians' queue (Q_{phys}) and nurses' final treatment queue (Q_{nurs}). The results consistently reduce the expected waiting time of patients in more flexible strategies. This is further confirmed by the similar behavior of the maximum expected waiting time over all replications, in columns *Qmax*. However, within this trend of reducing the average waiting time, the results are not very stable. Different set of scenarios in the optimization model result in different waiting time in the simulation model. For example with shifts of fixed length the waiting time varies between 79.94 minutes for the first set of scenarios and 87.28 for the second set. This can be explained by the approximations that were made and the fact that we only used a limited number of scenarios ($\sigma = 100$) to optimize.

Having schedules with very flexible shifts (in the range 1-24 hours) does not really improve the waiting time of patients compared to schedules with shifts in the range of 4-12 hours. Consequently, we note that modifying the starting times of the shifts is the factor that helps us most improve the expected waiting times, even with a constant total working hours capacity. As explained in Section 3, queues 1 and 2 are served by physicians and queue 3 is served by nurses. In the simulation model, the first two queues are grouped into a physician queue Q_{phys} .

Figures 5, 6 and 7 present example staffing levels obtained by optimizing the schedules following the different shift strategies. The reference schedule clearly misses the evening peak of arrivals while the other strategies closely match the arrival curve. We note that the solution with shifts varying from 4 to 12 hours and those varying from 1 to 24 hours are very close and match the ascending and descending slopes of the arrivals flow. These observations corroborate the results obtained in Tables 5 and 6.

6.3 Validation of the Robustness of the Schedules

In this set of simulation runs, we increment the number of patients arriving to the system and evaluate the schedules obtained through optimization again. In order to do so, we increment the average Poisson parameter for each period. This increase in the number of arrivals results in about 15% more patients per period. Since the schedules have not been optimized to perform under these conditions, the evaluation allows us to have an idea on the robustness of the schedules during epidemic periods.

The results of this set of tests are summarized in Tables 7 and 8. We notice from the results that

| Shifts | Wait (mins) | Q_{phys} (mins) | Q_{phys} max (mins) | Q_{nurs} (mins) | Q_{nurs} max (mins) |
|-------------|---------------|-------------------|-----------------------|-------------------|-----------------------|
| ref | 125,23 ± 6,22 | 99,24 ± 5,6 | 977,07 ± 43,24 | 25,99 ± 1,5 | 422,26 ± 23,08 |
| 8hrs (1) | 79,94 ± 4,85 | 54,66 ± 4,21 | 861,81 ± 46,89 | 25,28 ± 1,75 | 465,12 ± 38,88 |
| (2) | 87,28 ± 5,8 | 59,23 ± 5,13 | 910,35 ± 49,42 | 28,05 ± 1,87 | 499,15 ± 35,75 |
| (3) | 83,82 ± 7,61 | 58,73 ± 6,75 | 864,21 ± 52,76 | 25,09 ± 1,94 | 412,3 ± 39,43 |
| 4-12hrs (1) | 62,43 ± 4,3 | 48,91 ± 4,08 | 832,22 ± 43,19 | 13,52 ± 1,01 | 315,84 ± 32,67 |
| (2) | 63,13 ± 4,27 | 48,54 ± 3,94 | 836,35 ± 47,46 | 14,59 ± 0,99 | 374,4 ± 34,51 |
| (3) | 64,46 ± 3,7 | 50,72 ± 3,38 | 842,91 ± 34,87 | 13,74 ± 0,85 | 330,66 ± 28,85 |
| 1-24hrs (1) | 64,97 ± 4,58 | 51,06 ± 4,35 | 871,16 ± 45,72 | 13,91 ± 0,9 | 324 ± 30,8 |
| (2) | 64,55 ± 4,63 | 49,92 ± 4,16 | 855,5 ± 51,54 | 14,63 ± 1,02 | 369,36 ± 31,86 |
| (3) | 64,27 ± 4,16 | 48,18 ± 3,69 | 831,45 ± 42,44 | 16,09 ± 1,09 | 381,73 ± 28,08 |

Table 6: Simulation results for the weekend schedules

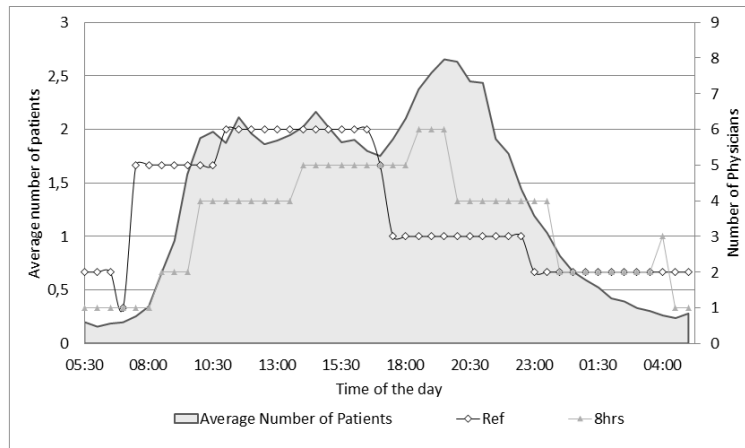


Figure 5: Resulting staffing levels with shifts fixed to 8 hours

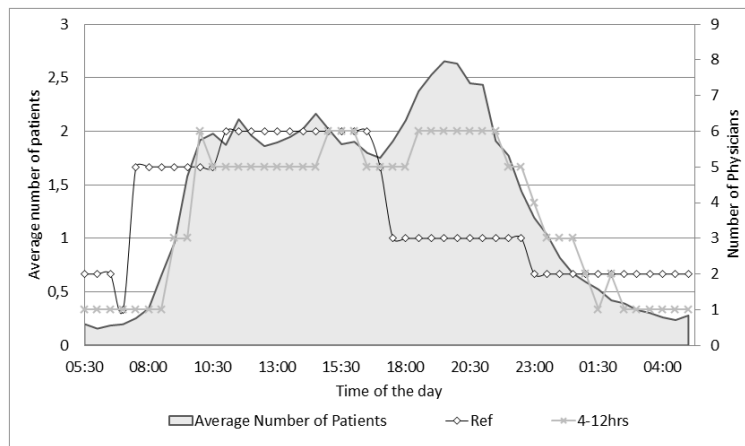


Figure 6: Resulting staffing levels for shifts varying from 4-12 hours

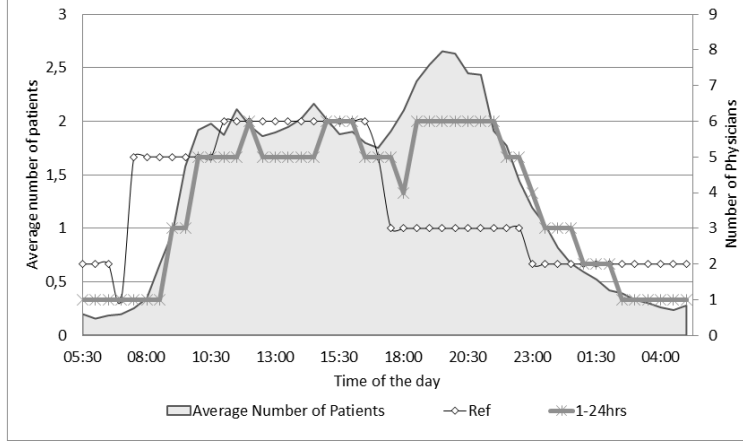


Figure 7: Resulting staffing levels for shifts varying from 1-12 hours

the expected waiting time increases proportionally to the results of Section 6.2 across the different solution instances. Consequently, the optimized schedules still outperform the reference schedule in unfavorable conditions. This coherence in the waiting times is still true for each queue individually but we see a disproportionate increase in the average maximum waiting times for schedules constructed with (1-24h) constraints. This increase is understandable as the (1-24h) constraints specifically exploit the discrete-time approximation of the optimization model. Schedules with (1-24h) constraints can be expected to perform poorly under different conditions than those they were optimized for.

| Shifts | Wait | Q_{phys} | $Q_{phys} \max$ | Q_{nurs} | $Q_{nurs} \max$ |
|------------|----------------|----------------|-----------------|--------------|-----------------|
| ref | 242,99 ± 13 | 199,2 ± 12,06 | 1323,22 ± 34,88 | 43,79 ± 2,61 | 559,87 ± 23,74 |
| 8hrs (1) | 191,1 ± 13,32 | 148,81 ± 13,11 | 1268,17 ± 45,18 | 42,29 ± 3,38 | 651,89 ± 45,34 |
| (2) | 191,32 ± 14,34 | 147,3 ± 13,53 | 1263,06 ± 44,4 | 44,02 ± 2,76 | 659,27 ± 45,17 |
| (3) | 190,79 ± 13,6 | 149,63 ± 13 | 1264,16 ± 45,57 | 41,16 ± 2,75 | 633,62 ± 44,73 |
| 4-12hrs(1) | 148,74 ± 10,86 | 125,47 ± 10,41 | 1218,3 ± 46,26 | 23,27 ± 1,85 | 457,79 ± 43,62 |
| (2) | 155,95 ± 12,04 | 133,56 ± 11,65 | 1258,86 ± 43,98 | 22,38 ± 1,43 | 494,64 ± 38,05 |
| (3) | 150,31 ± 10,85 | 128,77 ± 10,55 | 1242,86 ± 42,17 | 21,54 ± 1,59 | 477,88 ± 42,1 |
| 1-24hrs(1) | 155,76 ± 13,59 | 132,37 ± 13,22 | 1211,44 ± 49,83 | 23,39 ± 1,69 | 482,04 ± 41,28 |
| (2) | 146,73 ± 11,67 | 122,78 ± 11,02 | 1212,04 ± 49,64 | 23,94 ± 1,56 | 501,33 ± 30,19 |
| (3) | 142,07 ± 11,26 | 116,78 ± 10,99 | 1180,36 ± 48,45 | 25,29 ± 1,65 | 525,05 ± 34,35 |

Table 7: Simulation results for weekend schedules during an epidemic period

6.4 Week to Weekend Transition

One limitation of the previous results is that the cyclical schedules are evaluated independently for weekdays and weekends. In order to have simulation conditions that are closer to reality, the transition from a weekday schedule to a weekend schedule needs to be evaluated. Consequently, weekday and weekend schedules are concatenated and tested under both the average conditions for arrivals and the epidemic conditions.

The results in Tables 9 and 10 show that the transition from the week schedules to the weekend schedules can be done without real complications. The average waiting times observed when the concatenation of schedules is performed follow the results obtained previously. Indeed, the optimized schedules outperform the reference schedules under normal and epidemic conditions.

The results of the final set of tests are summarized in Figures 8 and 9. As we have mentioned, the

Table 8: Simulation results for week schedules during an epidemic period

| Shifts | Wait | Q_{phys} | Q_{phys} max | Q_{nurs} | Q_{nurs} max |
|------------|----------------|----------------|-----------------|--------------|----------------|
| ref | 142,07 ± 11,26 | 116,78 ± 10,99 | 1180,36 ± 48,45 | 25,29 ± 1,65 | 525,05 ± 34,35 |
| 8hrs (1) | 53,14 ± 3,17 | 22,53 ± 1,57 | 564,15 ± 42,14 | 30,61 ± 2,24 | 490,83 ± 37,81 |
| (2) | 51,62 ± 2,43 | 24,45 ± 1,45 | 580,12 ± 35,17 | 27,17 ± 1,65 | 502,22 ± 40,29 |
| (3) | 54,54 ± 3,44 | 21,44 ± 1,64 | 528,09 ± 42,14 | 33,1 ± 2,56 | 601,4 ± 36,75 |
| 4-12hrs(1) | 37,5 ± 2,62 | 12,8 ± 0,92 | 498,01 ± 30,82 | 24,7 ± 2,27 | 541,19 ± 37,59 |
| (2) | 35,11 ± 2,19 | 11,41 ± 0,75 | 417,48 ± 30,38 | 23,7 ± 1,82 | 524,46 ± 35,12 |
| (3) | 40,63 ± 2,4 | 13,58 ± 0,86 | 510,86 ± 28,89 | 27,05 ± 2 | 598,06 ± 36,17 |
| 1-24hrs(1) | 35,22 ± 1,95 | 11,94 ± 0,85 | 454,38 ± 30,59 | 23,27 ± 1,61 | 526,38 ± 33,41 |
| (2) | 35,39 ± 2,13 | 11,22 ± 0,73 | 414,83 ± 31,08 | 24,18 ± 1,77 | 549,97 ± 37,27 |
| (3) | 40,91 ± 2,22 | 13,32 ± 0,8 | 499,48 ± 26,06 | 27,59 ± 1,88 | 602,06 ± 32,71 |

| Shifts | Wait | Q_{phys} | Q_{phys} max | Q_{nurs} | Q_{nurs} max |
|-------------|--------------|--------------|----------------|--------------|----------------|
| ref | 88,59 ± 4,7 | 62,59 ± 3,63 | 871,71 ± 45,88 | 25,99 ± 1,92 | 430,6 ± 30,93 |
| 4-12hrs (1) | 38,52 ± 2,96 | 24 ± 2,4 | 652,18 ± 45,73 | 14,53 ± 1,07 | 367,98 ± 32,31 |
| (2) | 37,53 ± 2,84 | 22,7 ± 2,36 | 635,05 ± 46,81 | 14,83 ± 1,21 | 379,98 ± 34,39 |
| (3) | 38,48 ± 2,52 | 24,53 ± 2,27 | 654,31 ± 38,59 | 13,96 ± 0,97 | 384,82 ± 35,06 |

Table 9: Simulation results for schedules spanning an entire week

| Shifts | Wait | Q_{phys} | Q_{phys} max | Q_{nurs} | Q_{nurs} max |
|------------|---------------|---------------|-----------------|--------------|----------------|
| ref | 140,58 ± 6,63 | 102,25 ± 5,09 | 1087,38 ± 42,46 | 38,33 ± 2,52 | 561,25 ± 43,66 |
| 4-12hrs(1) | 69,31 ± 4,59 | 46,13 ± 3,74 | 941,59 ± 56,66 | 23,18 ± 1,9 | 504,82 ± 36,2 |
| (2) | 70,46 ± 4,61 | 43,86 ± 3,72 | 941,56 ± 49,76 | 26,6 ± 2,05 | 573,99 ± 38,44 |
| (3) | 71,09 ± 4,06 | 46,73 ± 3,32 | 974,81 ± 51,18 | 24,36 ± 2,12 | 508,22 ± 43,41 |

Table 10: Simulation results for schedules spanning an entire week during an epidemic period ϵ

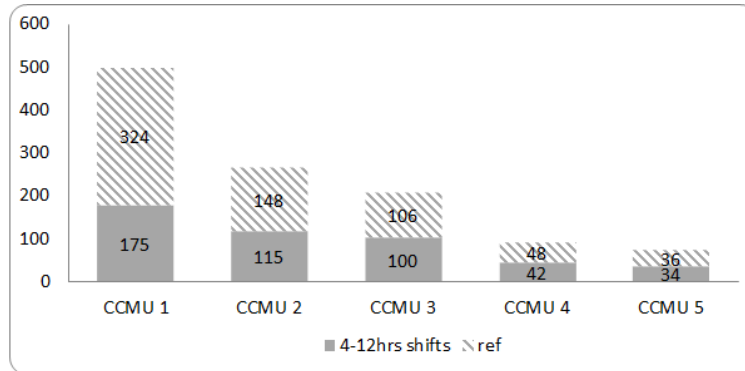


Figure 8: Expected sojourn time for the different acuity classes in normal conditions

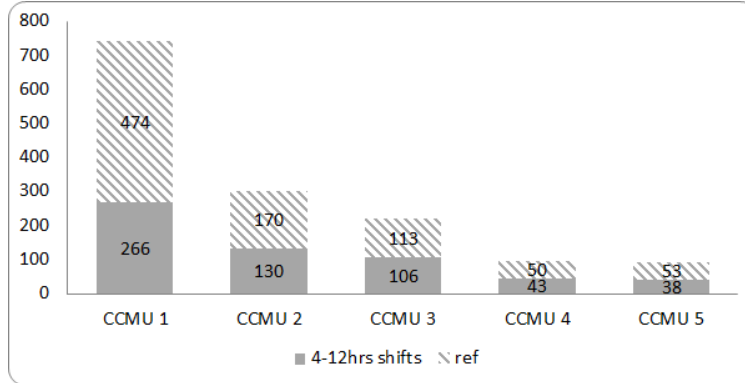


Figure 9: Expected sojourn time for the different acuity classes in epidemic conditions

schedules with shifts varying from 1 to 24 hours do not give a significant advantage over the schedules with shifts varying from 4 to 12 hours. Finally, Figures 8 and 9 illustrate the differences in sojourn time according to the CCMU of patients. Coherently, patients with higher priority spend less time in the system than those with lower priority.

7 Conclusion

We presented in this study a stochastic optimization formulation for an ED shift scheduling problem. Schedules of interest are computed once for some representative period such as winter or summer season or weekdays and weekends. During this period, the schedules are repeated daily. The formulation follows the flow of patients through the ED and takes into account the stochastic nature of arrivals and service times. The constraints used to follow the evolution of patients through the ED mirror those in stock management literature. The optimization model is solved using a deterministic Sample Average Approximation (SAA) method. This model assumes many simplifications in order to be tractable in a commercial linear programming solver. The solution schedules obtained by the optimization model are evaluated using a discrete event simulation model under various test conditions to verify the validity of the assumptions and the robustness of the solutions.

The simulation model used to evaluate the results has several improvements over the optimization model. Improvements such as patient prioritization and multiple tasks for resources (administrative tasks for physicians and biological exams for nurses) make the simulation model more realistic. The evaluation demonstrates that the assumptions made hold for optimizing the average waiting time. Nonetheless, the simulation model remains a simplification of the ED and hence lacks some elements that can be taken into account in future research. For example the service time of patients is approximated with an exponential distribution. Even if this approximation is very common in the literature, service times are usually not exponentially distributed. Hence, a possible extension to our work would be to test detailed patient service times to be collected from our partner hospital. Furthermore, excessive waiting time is only one aspect among many that define overcrowding in ED. Average waiting time is important and influences other criteria such as the maximum waiting time but other service quality criteria such as limited bed capacity contribute to overcrowding and should be considered in future research.

Real data from the UHC of Lille France are used to evaluate the quality and the performance of the schedules obtained. The proposed optimization approach allows to find promising cyclic schedules. The framework proposed is easily extensible and allows the integration of several strategical and operational constraints that are not possible using other methods.

8 acknowledgements

This research is partially supported by the French National Research Agency “ANR” under grant number ANR-11-TECS-0010 of project HOST.

References

- [1] Réduire les temps de passage aux Urgences. Tech. rep., Mission Nationale D’expertise et D’audit Hospitaliers. URL www.meah.sante.gouv.fr
- [2] Akkerman, R., Knip, M.: Reallocation of beds to reduce waiting time for cardiac surgery. *Health care management science* **7**(2), 119–26 (2004). URL <http://www.ncbi.nlm.nih.gov/pubmed/15152976>
- [3] Bruin, a.M., Bekker, R., Zanten, L., Koole, G.M.: Dimensioning hospital wards using the Erlang loss model. *Annals of Operations Research* **178**(1), 23–43 (2009). DOI 10.1007/s10479-009-0647-8. URL <http://link.springer.com/10.1007/s10479-009-0647-8>
- [4] Cochran, J.K., Roche, K.T.: A multi-class queuing network analysis methodology for improving hospital emergency department performance. *Computers & Operations Research* **36**(5), 1497–1512 (2009)
- [5] Edie, L.C.: Traffic delays at toll booths. *Operations Research* **2**(2), 107–138 (1954)
- [6] Ernst, A., Jiang, H., Krishnamoorthy, M., Sier, D.: Staff scheduling and rostering: review of applications, methods and models. *European Journal of Operational Research* **153**(1), 3–27 (2004). DOI 10.1016/S0377-2217(03)00095-X. URL <http://linkinghub.elsevier.com/retrieve/pii/S037722170300095X>
- [7] Green, L., Soares, J.a.: Using queueing theory to increase the effectiveness of emergency department provider staffing. *Academic Emergency Medicine* **13**(1), 61–68 (2006). DOI 10.1197/j.aem.2005.07.034. URL <http://www.ncbi.nlm.nih.gov/pubmed/16365329>
<http://onlinelibrary.wiley.com/doi/10.1197/j.aem.2005.07.034/abstract>
- [8] Ingolfsson, A., Amanul Haque, M., Umnikov, A.: Accounting for time-varying queueing effects in workforce scheduling. *European Journal of Operational Research* **139**(3), 585–597 (2002). DOI 10.1016/S0377-2217(01)00169-2. URL <http://linkinghub.elsevier.com/retrieve/pii/S0377221701001692>
- [9] Izady, N., Worthington, D.: Setting staffing requirements for time dependent queueing networks: The case of accident and emergency departments. *European Journal of Operational Research* **219**(3), 531–540 (2012). DOI 10.1016/j.ejor.2011.10.040. URL <http://linkinghub.elsevier.com/retrieve/pii/S0377221711009805>
- [10] Jaumard, B., Semet, F., Vovor, T.: A generalized linear programming model for nurse scheduling. *European journal of operational research* **2217**(97) (1998). URL <http://www.sciencedirect.com/science/article/pii/S0377221797003305>
- [11] Jennings, O.: Server staffing to meet time-varying demand. *Management Science* **42**(10), 1383–1394 (1996). URL <http://mansci.journal.informs.org/content/42/10/1383.short>
- [12] Jones, S., Thomas, A.: Forecasting daily patient volumes in the emergency department. *Academic Emergency Medicine* **15**(2), 159–170 (2008). DOI 10.1111/j.1553-2712.2007.00032.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/18275446> <http://onlinelibrary.wiley.com/doi/10.1111/j.1553-2712.2007.00032.x/full>
- [13] Jones, S.S., Evans, R.S., Allen, T.L., Thomas, A., Haug, P.J., Welch, S.J., Snow, G.L.: A multivariate time series approach to modeling and forecasting demand in the emergency department. *Journal of biomedical informatics* **42**(1), 123–39 (2009). DOI 10.1016/j.jbi.2008.05.003. URL <http://www.ncbi.nlm.nih.gov/pubmed/18571990>

- [14] Kleywegt, A.J., Shapiro, A., Homem-de Mello, T.: The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization* **12**(2), 479–502 (2002)
- [15] McCarthy, M.L., Zeger, S.L., Ding, R., Aronsky, D., Hoot, N.R., Kelen, G.D.: The challenge of predicting demand for emergency department services. *Academic Emergency Medicine* **15**(4), 337–46 (2008). DOI 10.1111/j.1553-2712.2008.00083.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/18370987>
- [16] Rogers, a.E., Hwang, W.T., Scott, L.D., Aiken, L.H., Dinges, D.F.: The Working Hours Of Hospital Staff Nurses And Patient Safety. *Health Affairs* **23**(4), 202–212 (2004). DOI 10.1377/hlthaff.23.4.202. URL <http://content.healthaffairs.org/cgi/doi/10.1377/hlthaff.23.4.202>
- [17] Schuur, J.D., Venkatesh, A.K.: The growing role of emergency departments in hospital admissions. *New England Journal of Medicine* **367**(5), 391–393 (2012)
- [18] Sinreich, D., Jabali, O.: Staggered work shifts: a way to downsize and restructure an emergency department workforce yet maintain current operational performance. *Health care management science* **10**(3), 293–308 (2007)
- [19] Sinreich, D., Jabali, O., Dellaert, N.P.: Reducing emergency department waiting times by adjusting work shifts considering patient visits to multiple care providers. *IIE Transactions* **44**(3), 163–180 (2012). DOI 10.1080/0740817X.2011.609875. URL <http://www.tandfonline.com/doi/abs/10.1080/0740817X.2011.609875>
- [20] Sven, O., Håkan, J., Jon, R., Katarina, G.: A systematic review of triage-related interventions to improve patient flow in emergency departments. *Scand J Trauma Resusc Emerg Med* **19**(43), 1–9 (2011). DOI 10.1186/1757-7241-19-43. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3152510&tool=pmcentrez&rendertype=abstract> <http://www.doaj.org/doi/func=abstract&id=819527>
- [21] Vile, J.: Time-dependent Stochastic Modelling for Predicting Demand and Scheduling of Emergency Medical Services. Ph.D. thesis, Cardiff University (2013). URL <http://orca.cf.ac.uk/43186/1/2013vilejphd.pdf>
- [22] Warner, D.: Scheduling nursing personnel according to nursing preference: A mathematical programming approach. *Operations Research* **24**(5), 842–856 (1976). URL <http://or.journal.informs.org/content/24/5/842.short>
- [23] Zeltyn, S., Marmor, Y.N., Mandelbaum, A., Carmeli, B., Greenshpan, O., Mesika, Y., Wasserkrug, S., Vortman, P., Shtub, A., Lauterman, T., et al.: Simulation-based models of emergency departments:: Operational, tactical, and strategic staffing. *ACM Transactions on Modeling and Computer Simulation (TOMACS)* **21**(4), 24 (2011)