CIRRELT-2020-10

A Data-Driven Approach to Include Availability of ICU Beds in the Planning of the Operating Room

Arnaud Augustin
Philippe Jouvet
Nadia Lahrichi
Andrea Lodi
Louis-Martin Rousseau

April 2020

Bureau de Montréal
Université de Montréal
C.P. 6128, succ. Centre-Ville
Montréal (Québec) H3C 3J7
Tél. : 1-514-343-7575
Télécopie : 1-514-343-7121

Bureau de Québec
Université Laval,
2325, rue de la Terrasse
Pavillon Palasis-Prince, local 2415
Québec (Québec) G1V 0A6
Tél. : 1-418-656-2073
Télécopie : 1-418-656-2624

A Data-Driven Approach to Include Availability of ICU Beds in the Planning of the Operating Room

Arnaud Augustin[1], Philippe Jouvet[2], Nadia Lahrichi[1,*], Andrea Lodi[1,3], Louis-Martin Rousseau[1]

1. Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Department of Mathematics and Industrial Engineering, Polytechnique Montréal
2. CHU Sainte-Justine, 3175, chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 1C5
3. Canada Excellence Research Chair in Data Science for Real-Time Decision-Making, Polytechnique Montréal, P.O. Box 6079, Station Centre-Ville, Montréal, Canada H3C 3A7

Abstract. In this paper, we propose a novel approach to deal with the integration of the cancellation probability due to congestion in the intensive care unit in the (long term) surgical case assignment problem. This problem consists of selecting patients (from the wait list) to be on the operating list of surgeons for a selected horizon, and assigning a day, an operating room, and a time block to each surgeon. We propose an approach that computes the probability of canceling cases on each day through the graph derived from a Markov Decision Process. This graph is then integrated into a Mixed Integer Programming model to optimized the monthly case assignment schedule. We evaluated the method on the practical case of the teaching hospital Sainte-Justine (CHUSJ) in Montreal. We show that prioritizing patients during this process only increases the quality of the schedule without decreasing the occupancy rate of the operating room. We also use probabilities that need to be discussed with senior management to decide on the acceptable risk to cancel an elective case.

Keywords: Operating room scheduling, ICU, data-driven approach

Acknowledgements. Partial funding for this project has been provided by CHU Sainte-Justine, the Canada excellence Research Chair in Data science for real-time decision-making and Hanalog, the Canada research chair in healthcare analytics and logisitcs. We gratefully acknowledge the support of the Fonds de recherche du Québec through their Infrastructure Grant.

---

* Corresponding author: nadia.lahrichi@polymtl.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2020

1 Introduction

The operating room (OR) is renowned for consuming about 40% of hospital budgets. Looking to improve its efficiency is a natural way to reduce cost and better use resources. Different criteria are widely used to measure the efficiency (Macario, 2006): start-time tardiness, case cancellation rate, turnover time, post-anesthesia care unit admission delays, etc. In this paper, we are interested in the case cancellation rate due to congestion in the ICU.

Addressing the OR management problem is challenging for both researchers and practitioners alike. An overview of the literature shows that, since the 2000s, the community has been very active on the subject because of both the complexity of the problem and the OR's potential economic impact. Classifications of the problems encountered in the OR are described in (Cardoen et al., 2010; Guerriero and Guido, 2011; Hulshof et al., 2012; Samudra et al., 2016). Briefly, the problems are usually addressed either at the strategic (e.g. master surgery planning),tactical (e.g. balancing hospital resources) or operational level (e.g. patient scheduling). Uncertainty and whether other services (upstream/ downstream) are included in the analysis are additional ways to classify problems in the OR.

In this paper, we focus on the integration of the intensive care unit (ICU) to the (long term) surgical case assignment problem (operational level). This problem consists of selecting patients (from the wait list) to be on the operating list of surgeons for a selected horizon, and assigning a day, an operating room, and a time block to each surgeon. In practice, each surgeon selects from his/her wait list a set of patients to operate on (usually one week in advance). The sequencing of the patients on the list may then be defined by the surgeon or management. The simple way to constitute the list is on a first in, first out basis, however this may not allow for effective use of the OR's time. We therefore need to determine the best mix each day to maximize utilization rate. In addition, knowing that some of the patients will require an ICU admission, we integrate the availability of beds in the surgical case assignment problem.

However, if ICU beds are not available, scheduled cases would need to be cancelled. As the ICU is known to be significantly impacted by uncertainty, i.e., the arrival of patients from the emergency room (where most of ICU patients come from) and the length of their stay.

To our knowledge, no study in the literature includes the ICU availability when solving the surgical case assignment problem (Fei et al., 2008; Mateus et al., 2018; Marques and Captivo, 2017). Most papers focus on demonstrating the impact of the OR on ICU efficiency and vice versa.

(Chow et al., 2011) show, for example, that a high bed occupancy rate in the ICU results in an increased burden on hospital staff, frequent cancellations of surgeries, and increased wait times. It is confirmed in (Fügener et al., 2016, 2014) that not considering the limited resources of intensive care and recovery beds during OR scheduling leads to a decrease in the level of service of these postoperative units. (Bowers, 2013) highlights the interdependence between resources in intensive care and the OR. The study shows that a balance between the number of surgeries performed and the number of beds in intensive care should reflect the relative cost of these two departments. (Cook et al., 2004) study the influence of parameters such as the distribution of arrivals, length of stay, resources in the ICU and the number of surgery rooms. Finally, in (Marmor et al., 2011), a simulation model is used to show the link between the level of service offered to patients and the bed occupancy rate. Several sources of variability are considered, such as the seasonality of operating theater programming or the length of stay in intensive care. More generally, the same type of impacts are observed in relation to the emergency room (McConnell et al., 2004). Increasing capacity in the ICU decreases the burden on the emergency room. Most of these papers use simulation to show the interaction between the OR and the ICU. This relation is then usually modeled by including either an indicator in the objective function (Vissers et al., 2005) or a hard constraint on the capacity of beds in the ICU while solving the OR scheduling problem (Santibanez et al., 2007; Pham and Klinkert, 2008; van Oostrum et al., 2008). In the latter case, this constraint is usually static and needs to be defined before solving the problem. In this paper, we propose a novel approach that integrates this constraint dynamically in the problem.

The contribution of this paper is thus to propose a scheduling approach for the OR that takes into account the stochastic availability of the ICU. We propose a novel approach that incorporates the graph derived from a Markov Decision Process into a Mixed Integer Programming model in order to compute the cancellation probability due to congestion in the ICU within the OR case assignment problem. We evaluate this approach on a practical case from the teaching hospital Sainte-Justine (CHUSJ), in Montreal, to analyze the data from the OR and the ICU.

This paper is structured as follows. Section 2 describes the case study of CHU Sainte-Justine. In Section 3, we introduce the mathematical model including the postoperative constraints. The approach for modeling the ICU is presented in Section 4. The results obtained will be presented and analyzed in Section 5 followed by the conclusion.

2 The practical case of CHU Sainte-Justine

CHU Sainte-Justine (CHUSJ) is a pediatric teaching hospital. The surgery department includes 85 surgeons covering more than 11 specialties (orthopedic surgery, plastic, general, etc.). Approximately 10,000 interventions are performed each year in the 14 available rooms. The objective of this section is, first, to describe the practical case of CHUSJ and, second, to understand how data compares with the literature. This is particularly relevant to generalize our approach and findings.

2.1 The OR at CHUSJ

We analyze data collected between January 1^{st}, 2013 and December 31^{st}, 2015. See Figure 1 for the distribution of interventions among specialties.

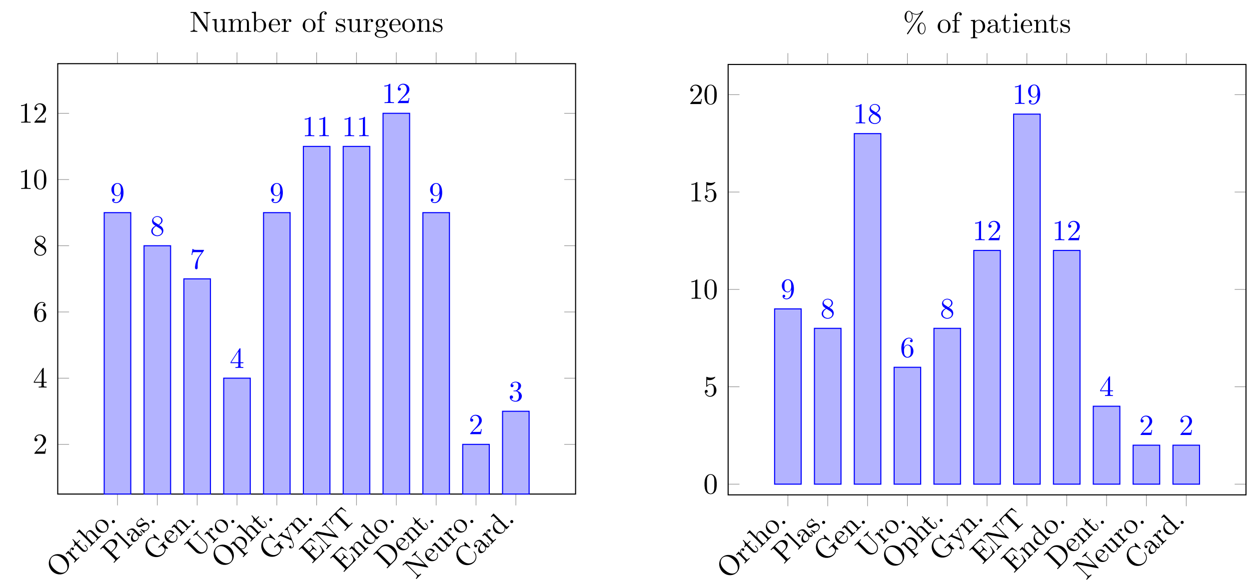
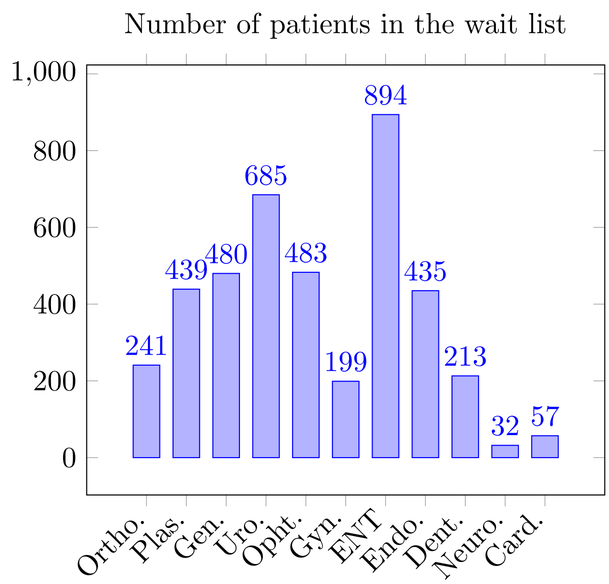


Figure 1: Distribution of surgeons and procedures per specialty

We observe that the two specialties with the highest number of interventions are general surgery and Ear, Nose and Throat (ENT) (almost 20% of the total load). On the other hand, neurosurgery and cardiology each represent 2% of the total number of interventions. Despite the smaller number, these latter patients have a significant impact on the use of upstream/downstream resources in the operating room. In particular, they usually require admittance to the ICU.

The OR schedule follows a master surgical schedule (MSS) that is updated monthly. It is usually based on government guidelines and how the hospital prioritizes specialties. Although it varies each month, it seems to be stable at CHUSJ. The distribution of time for each specialty varies between 50 hours (dentistry) and about 125 hours (for orthopaedic (ortho.) and general surgery). Cardiac surgery uses 62 hours approximately, plastic and endovascular (endo.) about 110h and the rest around 80-90 hours. This partially explains the small percentage of the total number of operated patients (see figure 1) as well as the large wait time (figure 2) that cardiac surgery represents. Less than half of the slots available in the month are open to interventions (130.5 slots available out of 280 in total). This is linked to the limited resources of the hospital (surgeons, nurses, anesthesiologists, hospital beds, etc.).

An analysis of the wait list of patients on May 2^{nd}, 2017 shows that over 4,000 patients were waiting for surgery. Figure 2 reports the number of patients waiting for each specialty and their average wait time (in days).



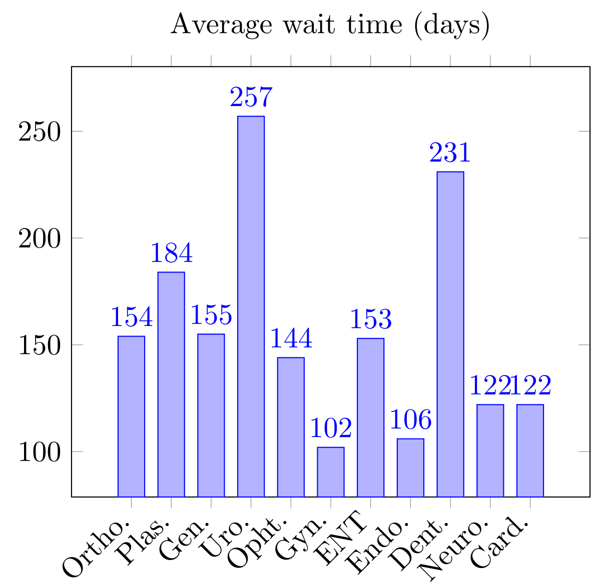


Figure 2: Number of patients in the wait list and average wait time - Wait list, May 2^{nd}, 2017

We observe that the highest number of patients on the wait list are from ENT and Urology. It is important to note that Urology is also one of the specialties with the fewest number of surgeons. On the contrary, many ENT surgeons are available and the number of interventions is the most important. If we now look at the average wait time, urology patients wait the most (approximately 8.5 months) followed by dentistry patients (around 8 months). The other specialties have relatively similar average wait times, around 5 months, except for gynecology for which the average wait is around 3 months.

We then analyze how the OR performs in terms of start time and end time. In theory, the OR opens at 8am and closes at 4pm. Figure 3 shows that most cases start on time, with very few

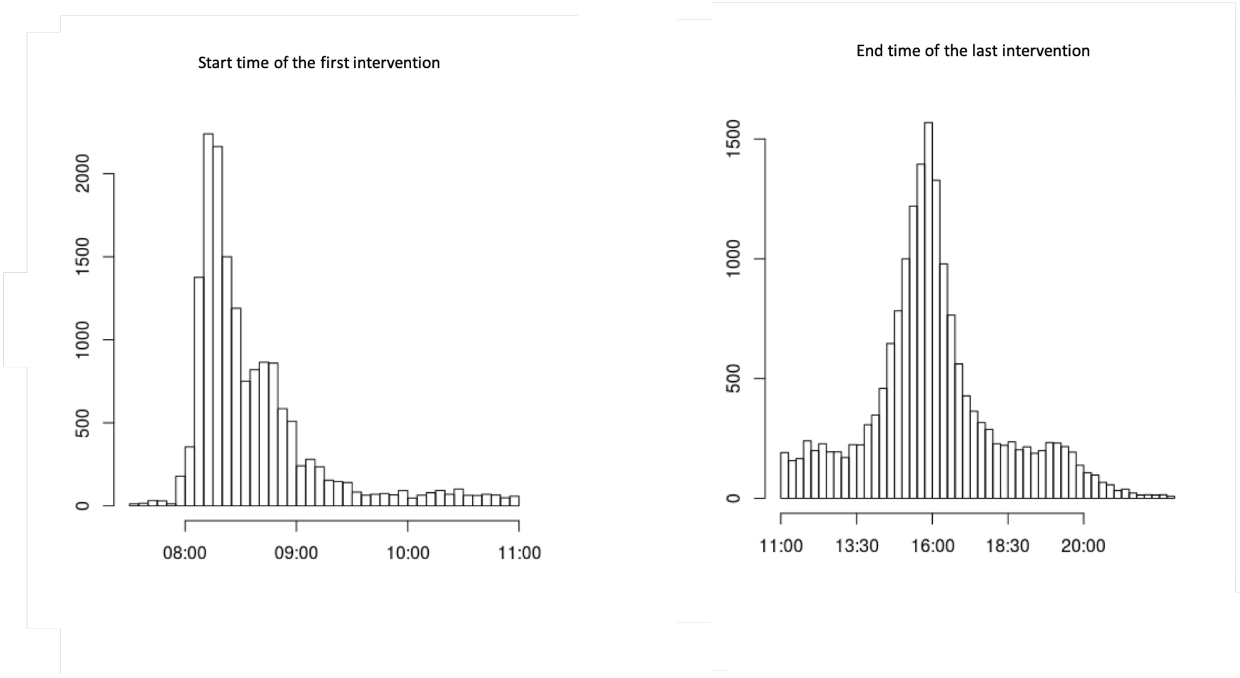


Figure 3: Start time and end time of rooms

rooms starting after 9am. While most rooms end at 4pm, some may end earlier or later.

Finally, an analysis of the duration of interventions shows it fits a Lognormal distribution.

2.2 The intensive care unit

Table 1 illustrates the number of patients requiring intensive care per specialty on the wait list as of January 26^{th}, 2017. It shows that 90% of cardiac patients (about 32 patients) needed an ICU recovery. Neurosurgery (5 patients) and orthopaedics (35 patients) require ICU in 13% of the cases. In total, 2% of elective patients require ICU recovery.

Table 1: Patients requiring postoperative ICU

	Ortho.	Plas.	Gen.	Uro.	Opht.	Gyn.	ENT	Endo.	Dent.	Neuro.	Card.	Total
ICU	35	2	5	3	0	0	3	0	1	5	32	86
Total	264	444	511	755	459	192	982	397	223	37	35	4299
Percentage (%)	13.3	0.5	1.0	0.4	0.0	0.0	0.3	0.0	0.4	13.5	91.4	2.0

Data available from the ICU, between April 6^{th}, 2014 and February 13^{rd}, 2017, shows that each year an average of 17% of the requests for surgery requiring an ICU admission are cancelled (see Table 2). This represents an average of 1.4 requests/day and 0.23 cancellations/day.

A finer analysis of the cancellations during the year shows that surgeries are more likely to be postponed or canceled during the winter (see Figure 4).

Table 2: Data on requests/cancellations per year

	2014-2015	2015-2016	2016-2017	Total
Number of requests	365	351	294	1010
Number of cancellations	75	52	47	174
Probability to cancel	21%	15%	16%	17%

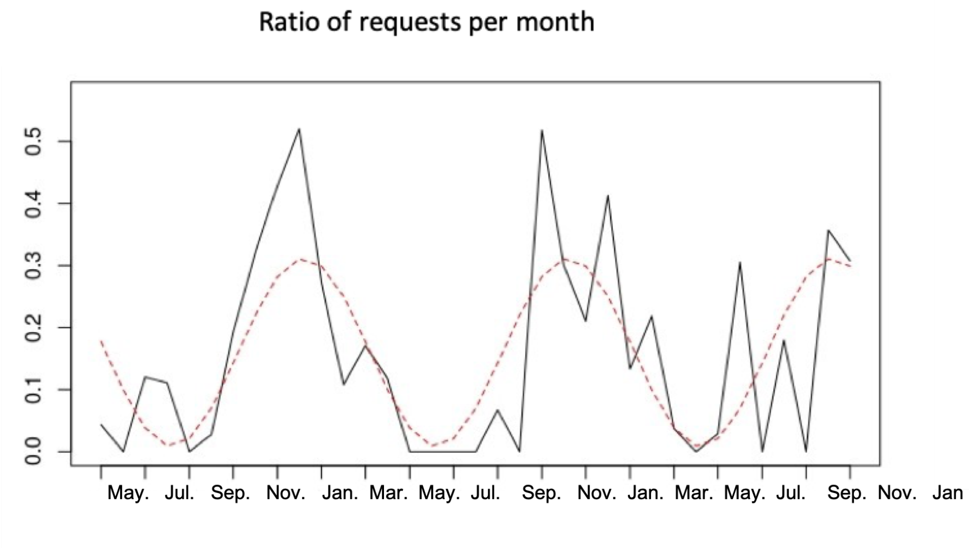


Figure 4: Ratio of surgeries with ICU admission canceled/postponed

Fitting the data (dotted line) shows an average of 1.4 cancellations and a maximum (minimum) reached in January (July). Based on this graph and a discussion with senior management from the hospital, we consider four periods in the year November-January, February-April, May-July and August-October. They may be aggregated into November-April and May-October.

Table 3: Data per period

<table>
<tr><td></td><td colspan="4">Periods</td></tr>
<tr><td></td><td>Nov. - Jan.</td><td>Feb. - Apr.</td><td>May - July.</td><td>Aug. - Oct</td></tr>
<tr><td rowspan="2">Number of requests</td><td>273</td><td>240</td><td>231</td><td>266</td></tr>
<tr><td colspan="2">513</td><td colspan="2">497</td></tr>
<tr><td rowspan="2">Number of cancellations</td><td>89</td><td>44</td><td>21</td><td>20</td></tr>
<tr><td colspan="2">133</td><td colspan="2">41</td></tr>
<tr><td rowspan="2">Probability to cancel</td><td>33%</td><td>18%</td><td>9%</td><td>8%</td></tr>
<tr><td colspan="2">26%</td><td colspan="2">8%</td></tr>
</table>

We clearly observe that a cancellation is more likely to occur in November-April than in the period May-October. A finer analysis per day (Figure 5) shows that demand is similar for both periods with a peak of almost 2 patients daily on Tuesdays and 1.5 patients on Thursdays. Since the requests are similar and the postponement rate varies significantly, we hypothesize that the ICU capacity for elective patients is greater between May-October. It also appears that the postponement rate is slightly different between November-January and February-April. A peak of postponements is observed on Wednesdays (green line) followed by a rate lower than average on the following day. This shows that there is enough capacity on that day.

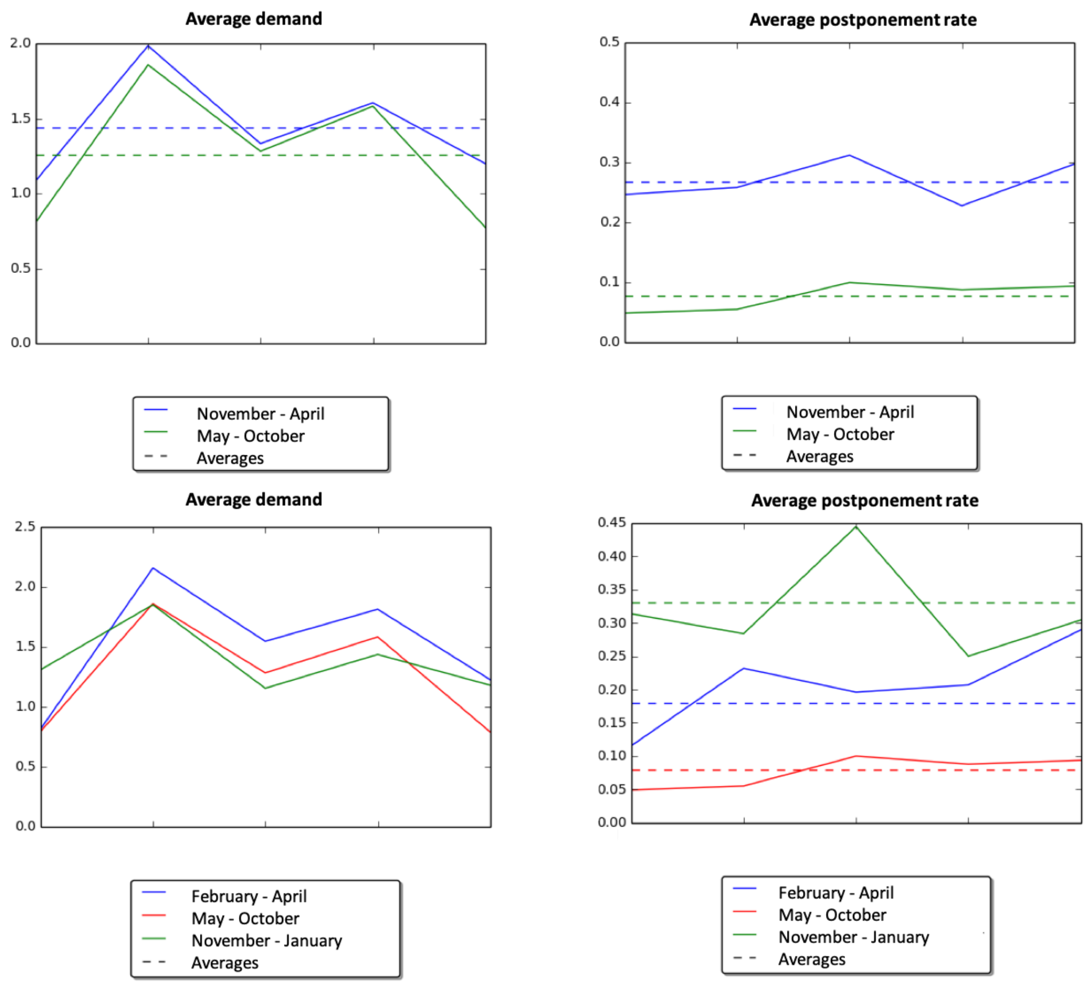
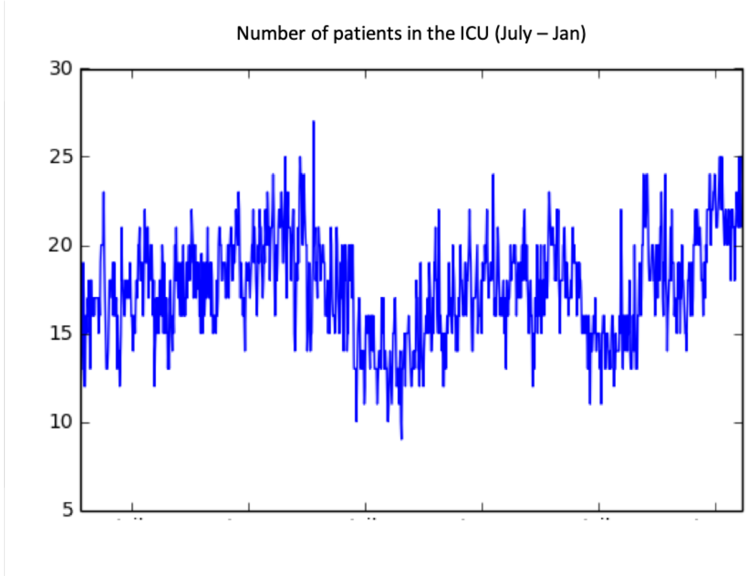


Figure 5: Requests and postponements over the periods

Figure 6 shows that the number of patients in the ICU (information collected everyday at 7am) is 17, on average.

The seasonality previously observed is confirmed in the figure. The number of patients decreases during summer time and capacity is higher between May-October. The analysis per day of the week (Figure 7) shows the same pattern regardless of the period of the year. The unit empties gradually during the week-end (no elective patients admitted during that time). Finally, we note a small

Figure 6: Occupancy in the ICU (July 2014 - January 2017)



variation on the number of patients in the ICU on Thursdays between November-April. This is probably related to postponement rate of elective patients on Wednesdays, observed in the same period.

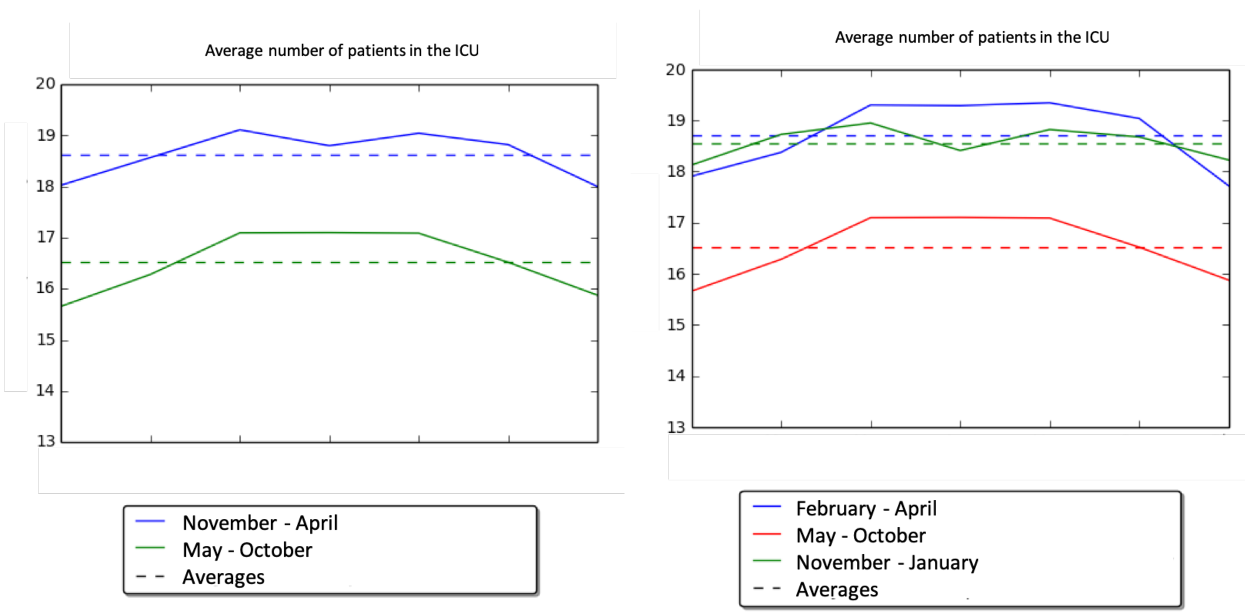


Figure 7: Average number of patients in the ICU per day

Figure 8 shows the number of patients (elective and non-elective) admitted in the ICU per month. Non-elective patients represent almost 75% of all patients (2,784 patients). These patients have priority in the ICU, and the variability in admissions is the major reason for elective case cancellation.

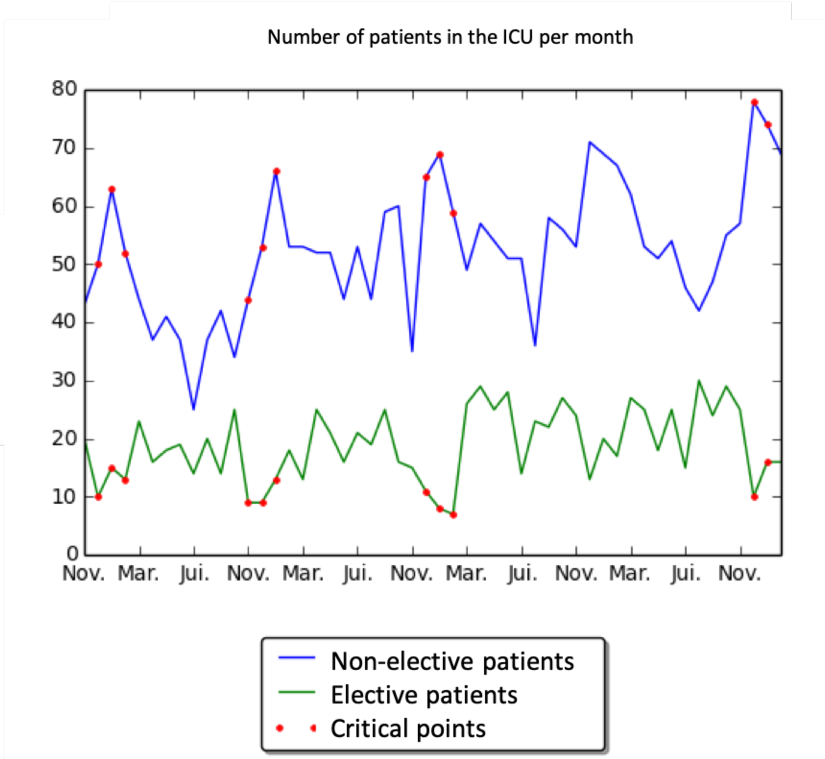


Figure 8: Elective and non-elective patients admission in the ICU (2012 - 2017)

The figure again confirms the observation that capacity decreases during Summer time. We also note a few outliers that may be explained by

- A decrease in demand;
- An increase in demand for non-elective cases;
- A decrease in the number of nurses in intensive care.

We clearly observe a correlation between the low number of admitted elective patients and the high number of non-elective patients (mainly emergency patients). We also observe a small increase of admission of non-elective patients (about one more per day). Finally, we observe a decrease in the number of non-elective admissions around July of each year.

A finer analysis using the Python library (https://facebookincubator.github.io/prophet/) shows the trends in the admission of emergency patients in the ICU. Figure 9 shows that since 2012, one additional emergency patient per day is admitted in the ICU. It also confirms the seasonality of the demand: November-April and May-October.

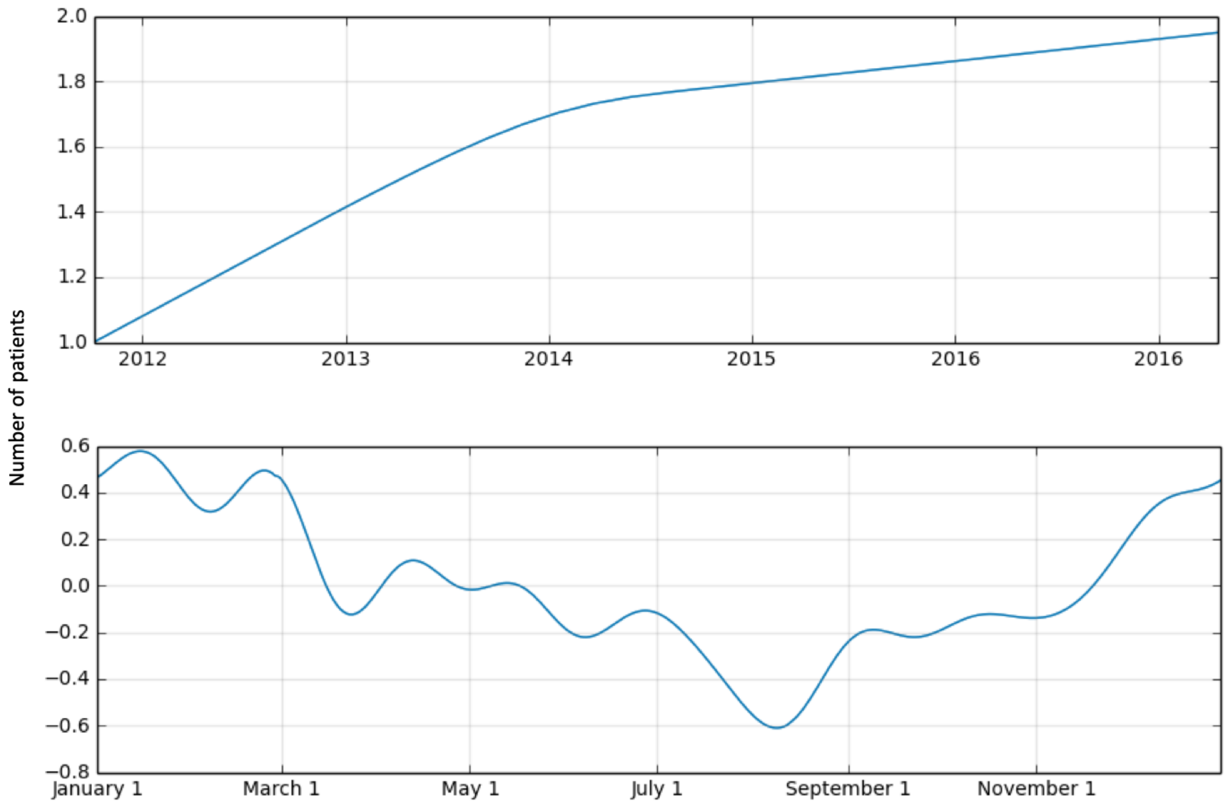


Figure 9: Trends in non-elective admissions in the ICU

Finally, the distribution of non-elective patients' arrival is very similar to the literature and follows a Poisson distribution.

An average of 2.5 discharges from the ICU is observed every day (see Figure 10). Most of them occur between Tuesdays and Fridays, regardless of the period considered.

2.3 Discussion

A recent publication (Leeftink and Hans (2018)) on benchmark instances for case mix classification and surgery scheduling confirms that our data from CHUSJ is aligned with the one in the literature:

- the 3-parameter lognormal distribution best fits surgery duration distribution;
- the usual opening hours of each OR is 8h;
- the instances used 11 specialties (17 case mix profiles), between 5 and 40 ORs, a load parameter that determines the expected surgery workload set between 0.8 and 1.2.

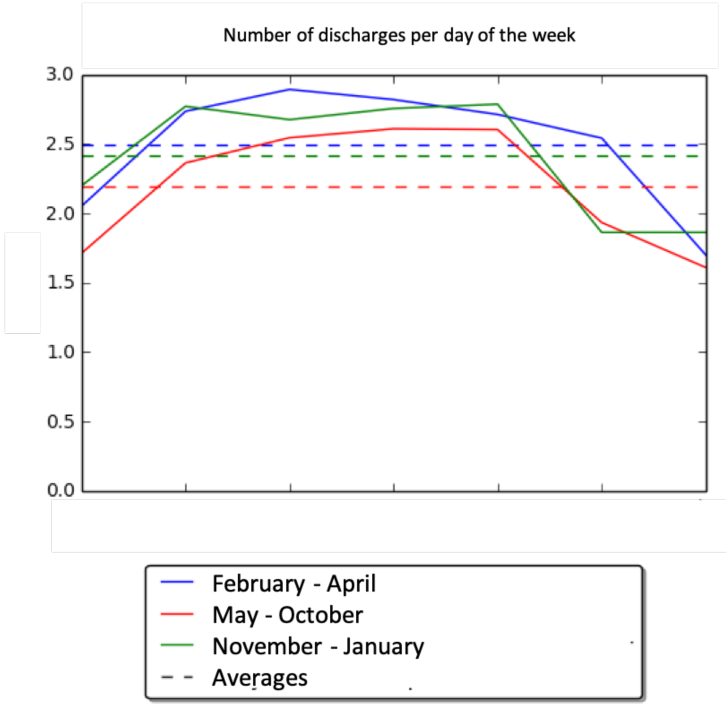


Figure 10: Distribution of discharges per day of the week

Therefore, our method and results are not instance dependent.

Finally, in most papers that integrate ICU in the literature, the number of beds available each day is very similar to what we observed in CHUSJ: 1 bed available for surgery patients in (Pham and Klinkert (2008)) and an average of 2 beds in (Santibanez et al. (2007)). Note that in the latter case, they refer to special care units rather than intensive care units.

3 The surgical case assignment problem

Our objective is to solve the surgical case assignment at CHUSJ. Namely, we want to maximize room occupancy by prioritizing the oldest patients on the wait list. We define four priority groups of the same size. We need to select patients to operate during the horizon, determine the slots for intervention of these patients, and assign surgeons to available slots. Currently, CHUSJ uses a fixed master surgical schedule. This MSS is generated monthly, and indicates the specialty and the room assigned to it each day. Although most of the slots of the OR schedule may already be assigned to a surgeon, some slots remain available. In the latter case, one needs to allocate the surgeon in order to maximize the objective function.

Note that a patient can only be operated by the pre-assigned surgeon.

The parameters and variables used in our model are summarized in Table 4.

Table 4: Notation used in the model

	Notation	Definition
Sets	T	Number of days in the horizon (28 days)
	C	Set of surgeons
	S	Set of specialties
	I	Set of patients
	K	Set of rooms
Parameters	Ω_{tk}	Availability in minutes of room k on day t
	α	Load parameter of operating rooms
	d_i	Duration of intervention for patient i
	γ_i	Priority group of patient i
	q_{cs}	1 if surgeon c is assigned to specialty s and 0 otherwise
	b_{ic}	1 if patient i is assigned to surgeon c and 0 otherwise
	mss_{tk}	c if surgeon c is assigned to room k on day t and 0 otherwise
	e_{stk}	1 if room k is assigned to specialty s on day t and 0 otherwise
Variables	r_{ctk}	1 if surgeon c is assigned to room k on day t and 0 otherwise
	x_{itk}	1 if patient i is assigned to room k on day t and 0 otherwise

The model is the following:

$$\max_{r,x} \qquad \sum_{i\in I}\sum_{t\in T}\sum_{k\in K} x_{itk} d_i \gamma_i \tag{1}$$

subject to:

$$\sum_{c\in C} r_{ctk} \leq 1 \qquad \forall t \in T, k \in K \tag{2}$$

$$r_{ctk} \leq \sum_{s\in S} q_{cs} e_{stk} \qquad \forall c \in C, t \in T, k \in K \tag{3}$$

$$x_{itk} \leq r_{ctk} \qquad \forall c \in C, t \in T, k \in K, i \in I | b_{ic} = 1 \tag{4}$$

$$\sum_{t\in T}\sum_{k\in K} x_{itk} \leq 1 \qquad \forall i \in I \tag{5}$$

$$\sum_{i\in I} x_{itk} d_i \leq \alpha \Omega_{tk} \qquad \forall t \in T, k \in K \tag{6}$$

$$r_{ctk} = 1 \qquad \forall c \in C, t \in T, k \in K | mss_{tk} = c \tag{7}$$

$$r_{ctk} \in \{0, 1\} \qquad \forall c \in C, t \in T, k \in K \tag{8}$$

$$x_{itk} \in \{0, 1\} \qquad \forall i \in I, t \in T, k \in K. \tag{9}$$

The objective function (1) maximizes room occupancy while prioritizing the oldest patients on the wait list. The score of each patient is equal to $d_i\gamma_i$. As in Agnetis et al. (2012), considering the duration at this point ensures we are not only scheduling short cases.

Constraints (2) ensure that up to one surgeon can be assigned to a room each day and constraints (3) that surgeon c can only be assigned to room k on day t if his/her specialty is already scheduled in the MSS. Constraints (4) link a patient i to a slot only if his/her surgeon is also assigned to it. Constraints (5) ensure each patient is processed, at most, once in the horizon. Capacity duration of rooms is enforced by constraints (6). No extra time is allowed if $\alpha \leq 1$. Constraints (7) impose the assignment of surgeons already assigned to slots in the MSS. Finally, constraints (8-9) state that variables are binary.

Next section describes how to include the availability in the ICU to this problem.

4 Modeling the availability in the ICU

Better anticipating the admission capacity in the ICU at the time of scheduling the OR will limit the risk of last minute cancellations. As mentioned earlier, it is very rare that more than 2 elective patients are admitted in the ICU. Instead, 75% of capacity is occupied by emergency patients.

We propose an approach based on graphical optimization models. We first model the system using Markov Decision Processes, using state and action variables. A state is defined by explanatory variables of the number of cancellations and, in each state, we determine (action) how many patients to schedule for the next day. We first define the explanatory variables of these cancellations in order to present the states.

As discussed in Section 2, two seasonal factors influence the probability of an elective admission: the month of the year and the day of the week. We propose a third factor which is the number of patients canceled the day before. For example, if two patients are canceled on day t, the probability of admitting an elective patient is higher on $t+1$. Each state is therefore defined by three parameters: the season, the day of the week, and the number of cancellations.

We can derive the number of patients requiring an ICU admission once the surgical case mix is obtained each day. Thus, for each schedule, we build a graph, where the capacity of the edges is defined as the probability of transition from one state to another in the MDP. The intuition is that given an initial state (for example "Tuesday") and a request for one ICU bed, the flow will propagate through the different possible states in this graph. Each state is associated with a number of cancellations. Flow conservation constraints are enforced. The probability of reaching each state of the network at a time t will help quantify the risk of canceling the number of scheduled cases with ICU admission. Note that these probability transitions are data-driven and thus may depend on the season.

Let's introduce a simple example with six days (Monday to Friday and one day to model the week-ends), and a maximum number of patients that can be canceled each day equal to 2. The maximum number of patients to schedule per day is two as well (derived from the decision variable in case mix assignment optimization model). Each state is defined by the day of the week and the number of cancellations. The transition probabilities are data-driven. In this case, we will use both the two periods of the year derived from the analysis in Section 2.2. For each day, 18 possible states are possible (6 days and 0 to 2 cancellations) and we can decide to schedule 0, 1 or 2 patients for the next day. This decision process is graphically illustrated in Figure 11.

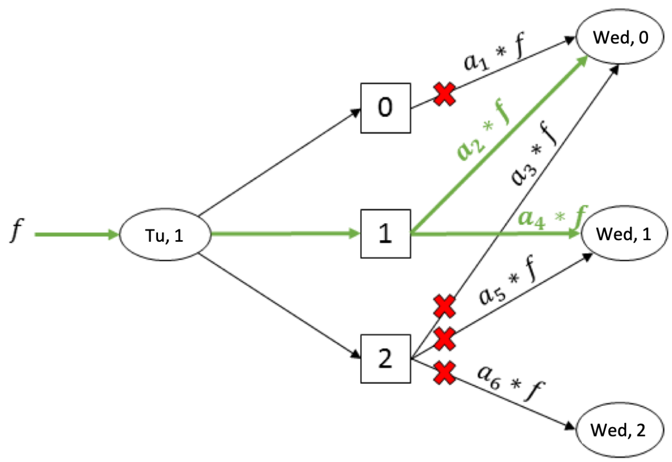


Figure 11: Illustration of a node in the graphical model

In this example, we are in the state (Tu, 1) which refers to Tuesday, with one elective patient being canceled. The decision for the next day will lead to a move to the state (Wed,0), i.e., "Wednesday" with 0 cancellations, or (Wed,1), (Wed,2). If the action is to place **1 request** for an ICU admission for the next day, this will lead to 0 cancellations (i.e. will be accepted) with a probability a_2 and to 1 cancellation with a probability a_4. Note that only six outgoing arcs are represented (three of the nine possibilities have a probability equal to 0, refer for example to the arc corresponding to 1 request and leading to 2 cancellations which is impossible in practice). These 6 arcs represent:

- 0 request leads to 0 cancellation with probability a_1 (equal to one);
- 1 request leads to 0 or 1 cancellation with probability a_2 and a_4;

- 2 requests lead to 0, 1 or 2 cancellations with probability a_3, a_5 and a_6.

The capacity a of each of these arcs is equal to the probability of having K cancellations (indicated in the arrival node) **given that we are in the state corresponding to the current node**, here (Tu, 1). Since the request is equal to 1 in this case, all flow outgoing from 0 and 2 is null (illustrated with a red cross). Finally, the flow f **needs to** spread across the remaining edges in proportion to their capacity, here a_2 and a_4, which makes this a particular case of network flow. The conservation of the flow is maintained with equation $a_2 + a_4 = 1$.

The complete network consists of the 18 nodes, corresponding to the states, repeated each day, forming different successive vertical layers. The number of layers is then equal to the number of days in the horizon and the flow entering each layer is equal to 1. We have

- L : Maximum number of elective patients scheduled per day;
- P : Maximum number of cancelled patients per day;
- $a_{tp_1\ell p_2}$: Probability to cancel p_2 patients given the state (day, cancellation) (t,p_1) and that ℓ patients are scheduled for the following day;
- $W_{t\ell} = \begin{cases} 1 \text{ if } \ell \text{ patients are scheduled on day } t \\ 0 \text{ otherwise} \end{cases}$
- and the variables $z_{tp_1\ell p_2}$ = flow (of probability) in arc between nodes (t, p_1) and $(t+1, p_2)$ associated with demand ℓ. The capacity of this arc is equal to $a_{tp_1\ell p_2}$.

Figure 13 in Appendix illustrates a horizon of 6 days and the network layers are represented from left to right. The first and last layers contain only one node and the others three nodes. We only represent the arcs for which the flow may be positive. The value shown in square brackets for each state represents the flow passing through this node. Note that the flow through each layer is equal to 1, which respects the flow conservation constraint. All probabilities are displayed.

We now introduce the linear model allowing to obtain the flow z. Let's suppose that the OR schedule is available for the week and is an input parameter of problem W. Then, $z_{tp_1\ell p_2}$ is the flow (of probabilities) propagating through the edge connecting nodes (t, p_1) and $(t+1, p_2)$ and associated with the request ℓ (capacity of the arc is $a_{tp_1\ell p_2}$).

The model is as follows:

$$\max_{z} \quad 0 \tag{10}$$

subject to:

$$0 \leq z_{tp_1\ell p_2} \leq a_{tp_1\ell p_2} \qquad \forall t \in T, (p_1, p_2) \in P, \ell \in L \tag{11}$$

$$z_{tp_1\ell p_2} \leq a_{tp_1\ell p_2} \sum_{p_0,\ell_0} z_{(t-1)p_0\ell_0 p_1} \qquad \forall t \in T, (p_1, p_2) \in P, \ell \in L, t \geq 1 \tag{12}$$

$$\sum_{p_1,p_2} z_{tp_1\ell p_2} \leq W_{t\ell} \qquad \forall t \in T, \ell \in L \tag{13}$$

$$\sum_{p_1,\ell,p_2} z_{tp_1\ell p_2} = 1 \qquad \forall t \in T, t \geq 1 \tag{14}$$

$$\sum_{\ell,p_2} z_{00\ell p_2} = 1. \tag{15}$$

The objective function (10) is to find a feasible flow (of value 1). Note that the flow is unique. Constraints (11) and (12) are capacity constraints that ensure the outgoing flow of a node is spread on the edges according the probabilities of the action one can take at each state. Constraints (13) define existing arcs given ℓ, the number of patients scheduled on day t. Finally, the flow from the initial node ($t = 0$, $p_1 = 0$) is equal to 1 (constraints (15)) and constraints (14) are flow conservation constraints (flow of 1 between each layer).

This model gives us, for a given OR schedule, a flow through each node representing the probability to transit through this state. These flows thus indicate, for each day, the probability of having 0, 1 or 2 cancellations when the OR schedule is defined by parameter W.

Senior management plays an active role at this stage to define the acceptable risk for canceling 1 or 2 cases per day. This upper bound is then included in the initial model to schedule the OR.

When merging both the graphical model to obtain the flow and the case mix assignment model, $W_{t\ell}$ is no longer a parameter but the result of the scheduling problem (x variables). Let g be a function that returns the number of intensive care patients scheduled on day t in the schedule (defined by the x variable). Constraints (16-17) link variables x and W

$$\sum_{\ell} \ell W_{t\ell} = g(x, t) \qquad \forall t \in T \tag{16}$$

$$\sum_{\ell} W_{t\ell} = 1 \qquad \forall t \in T. \tag{17}$$

We also introduce a binary parameter ω_i that indicates whether a patient requires an ICU admission (equals 1) or not (equals 0). The model finally becomes

$$\max_{r,x,z,W} \qquad \sum_{i\in I}\sum_{t\in T}\sum_{k\in K} x_{itk} d_i \gamma_i \tag{18}$$

subject to:

$$(2) - (9)$$

$$0 \leq z_{tp_1\ell p_2} \leq a_{tp_1\ell p_2} \qquad \forall t \in T, (p_1, p_2) \in P, \ell \in L \tag{19}$$

$$\sum_{p_1,p_2} z_{tp_1\ell p_2} \leq W_{t\ell} \qquad \forall t \in T, \ell \in L \tag{20}$$

$$\sum_{l,p_2} z_{00\ell p_2} = 1 \tag{21}$$

$$\sum_{p_1,l,p_2} z_{tp_1\ell p_2} = 1 \qquad \forall t \in T, t \geq 1 \tag{22}$$

$$z_{tp_1\ell p_2} \leq a_{tp_1\ell p_2} \sum_{p_0,l_0} z_{(t-1)p_0 l_0 p_1} \qquad \forall t \in T, (p_1, p_2) \in P, \ell \in L, t \geq 1 \tag{23}$$

$$\sum_{\ell} \ell W_{t\ell} = \sum_{i,k} \omega_i x_{itk} \qquad \forall t \in T \tag{24}$$

$$\sum_{\ell} W_{t\ell} = 1 \qquad \forall t \in T \tag{25}$$

$$W_{t\ell} \in \{0, 1\} \qquad \forall t \in T, \ell \in L \tag{26}$$

Constraints (19-26) relate to the distribution of patients requiring ICU during the week in order to limit the risk of cancellation. This block of constraints gives us, in particular, the variable z which quantifies this risk. The last step is therefore to add the two constraints

$$\sum_{p_1,\ell} z_{t-1,p_1,\ell,1} \leq u_1 \qquad \forall t \in T, t \geq 1 \tag{27}$$

$$\sum_{p_1,\ell} z_{t-1,p_1,\ell,2} \leq u_2 \qquad \forall t \in T, t \geq 1, \tag{28}$$

where u_1 and u_2 are the maximum probabilities of getting 1 and 2 cancellations, respectively. Recall that the involvement of senior management to determine these values is essential.

In the next section, we discuss the results based on the practical case of CHU Sainte-Justine.

5 Experiments and results

To evaluate our scheduling approach, we use six performance measures that focus on the efficiency of the operating room, namely

1. **Occupancy rate (ρ):** ratio of the sum of the duration of the programmed interventions to the total duration of programming available on the horizon.

2. **Group priority distribution:** This distribution indicates the number of scheduled patients in each group. For example, $\begin{pmatrix} Q_4 \\ Q_3 \\ Q_2 \\ Q_1 \end{pmatrix} = \begin{pmatrix} 89 \\ 71 \\ 13 \\ 0 \end{pmatrix}$ indicates that 89 in the fourth group (with the highest wait time), 71 in the third group, 13 in the second and 0 patients in the group of the most recent patients are scheduled.

3. **Average Wait Time (μ_{wait}):** The average wait time is the average wait time for patients scheduled on the horizon. A high wait time means that patients scheduled are "old" patients from the wait list.

4. **Standard deviation of wait time (σ_{wait}):** The smaller this standard deviation, the more patients are scheduled among the oldest patients.

5. **Relative standard deviation (RSD_{wait}):** This coefficient represents the ratio between the standard deviation and the average of the wait times. It is a standardized measure that shows the extent of variability to the mean. When the value is close to 0, it reflects the absence of variability in the data. This ratio will help compare the selection of patients between models in an objective way.

6. **Potential number of patients canceled (# cancellations):** This number is obtained after running a simulation (5000 replications) that uses the transition probabilities between states.

We compare these metrics on different alternatives to include the ICU availability. Parameters u_1 and u_2 allow to quantify the risk of cancellation for each day to integrate in the case assignment problem. They correspond respectively to the maximum probability of having 1 and 2 cancellations. These probabilities need to be discussed with senior management to decide on the acceptable risk to cancel an elective case. We have tested the following three alternatives:

- Static constraint to limit the number of elective patients requiring ICU to two (≤ 2);

- Dynamic constraint with no intervention from the manager to limit the risk of cancellation, $u_1 = 1, u_2 = 1$;
- Dynamic constraint with different levels of intervention from the manager to limit the risk: $u_1 = 0.5; u_2 = 0.25$;(the risk of canceling one patient is less than 550% and that of canceling two patients is less than 25%); $u_1 = 0.3; u_2 = 0.15$; $u_1 = 0.2, u_2 = 0.05$; $u_1 = 0; u_2 = 0$. Decreasing values of u_ℓ correspond to increasing risk aversion.

Finally, to quantify the benefit of the dynamic constraint, we run a simulation (5000 replications) to estimate the number of patients that may be canceled for each scenario.

All experiments are conducted on two real instances. For each instance, we use the wait list of patients (accessed once a month from the hospital data warehouse), the MSS of the same month, the estimation of the duration of interventions, the list of available surgeons and the data-driven transition probabilities to cancel cases requiring ICU. We use *high*, *medium* and *low* transition probabilities.

Note that since we do not consider turnover (time between two cases) in the model, we set α, the load parameter, to $\frac{7}{8}$. This represents 420 minutes for a complete day and 210 minutes for a half-day. This value is based on the average turnover observed in the OR of CHUSJ. About 50% of the MSS is not fixed (i.e. we schedule 50% of the rooms: specialty, surgeon and cases).

All results are obtained in approximately 10 seconds with an integrality GAP of less than 1%.

Tables 5 and 6 show the performance measures for each instance. The main take message from these results is that including availability of the ICU does not negatively impact the efficiency of the OR. Cases are scheduled differently to reduce risk of cancellation. This observation is valid for all scenarios.

Table 5: Average performance measures - Instance 1

				P_{medium}					P_{low}					P_{high}				
	No prio	Off	Static	1-1	0.5-0.25	0.3-0.1	0.2-0.05	0-0	1-1	0.5-0.25	0.3-0.1	0.2-0.05	0-0	1-1	0.5-0.25	0.3-0.1	0.2-0.05	0-0
ρ	98%	97%	97%	96%	97%	96%	97%	91%	96%	97%	97%	97%	93%	97%	97%	97%	96%	93%
Q_4	137	412	413	423	421	418	410	396	426	417	408	407	391	423	417	413	412	391
Q_3	129	158	156	151	162	166	164	174	161	161	160	167	164	145	161	165	170	164
Q_2	142	46	46	43	44	44	48	42	42	36	51	46	45	46	43	46	48	45
Q_1	170	29	32	34	33	30	32	32	36	35	31	33	37	35	31	35	34	37
# of patients	578	645	647	651	660	658	654	644	665	649	650	653	637	649	652	659	664	637
μ_{wait}	133	237	235	235	234	233	235	243	236	235	236	234	238	241	235	234	234	238
σ_{wait}	126	148	120	118	117	118	119	128	115	117	119	122	122	118	120	116	118	122
RSD_{wait}	0.95	0.62	0.51	0.50	0.50	0.51	0.50	0.53	0.49	0.50	0.50	0.52	0.51	0.50	0.51	0.50	0.50	0.551

The first column shows the results when no prioritization of patients is used (No prio), the second one includes the prioritization (Off). The availability in the ICU is not considered. The third column includes the availability of the ICU as a static constraint (Static). All the following

Table 6: Average performance measures - Instance 2

	No prio	Off	Static	P_{medium}					P_{low}					P_{high}				
				1-1	0.5-0.25	0.3-0.1	0.2-0.05	0-0	1-1	0.5-0.25	0.3-0.1	0.2-0.05	0-0	1-1	0.5-0.25	0.3-0.1	0.2-0.05	0-0
ρ	97%	96%	95%	96%	95%	96%	95%	91%	96%	96%	96%	95%	93%	95%	95%	95%	95%	93%
Q_4	147	452	455	474	4655	468	462	462	461	4855	454	471	465	474	460	454	456	465
Q_3	110	101	100	99	94	106	100	101	103	105	98	105	97	102	103	97	100	97
Q_2	149	30	35	32	34	36	34	31	32	34	32	34	28	33	35	33	33	31
Q_1	164	14	13	12	13	13	13	11	12	14	17	12	11	15	16	11	10	11
# of patients	570	597	603	617	606	619	609	602	608	638	605	622	604	624	614	595	599	604
μ_{wait}	143	271	262	266	262	261	260	261	264	262	256	261	257	261	259	262	263	257
σ_{wait}	141	142	119	119	117	117	117	117	122	113	117	114	121	124	117	120	122	121
RSD_{wait}	0.99	0.52	0.45	0.43	0.45	0.45	0.45	0.45	0.45	0.43	0.48	0.44	0.47	0.43	0.45	0.46	0.46	0.47

columns include the prioritization of patients and the probabilities to cancel cases.

When comparing the first two columns, we first observe that introducing the prioritization of patients does not impact the occupancy rate of the rooms (over 92%, on average, in all cases except for cardiac surgery). We also observe that we can improve the number of patients scheduled by 9% (about 50 patients for each instance). Regarding μ_{wait}, the average wait time, we observe that the scheduled patients are about twice as old with prioritization. In addition, we observe a significant decrease of δ_{wait} the coefficient of variation for all specialties (average of 41%) except for “Neurology” and “Cardiology”. Also, the **priority group distribution** decreases in most cases, except again for Cardiac and Neurosurgery. Although prioritization slightly improves the situation for these specialties, the scheduled patients remain divided almost equally between the different groups. This is due to the fact that there is very little flexibility in terms of duration of interventions (all usually long). Finally, it should be noted that, of all the specialties, about 65% of the patients scheduled are from the group of the oldest patients and that 90% come from half of the oldest patients. Figure 12 shows the distribution of the number of patients in each quartile Q_i per specialty.

When considering availability of the ICU (looking at the remaining columns), same observations stand. There is no negative impact on the efficiency of the OR. Utilisation room (ρ) is very stable (approximately 96%) and the only significant difference is observed for scenario $u_1 = u_2 = 0$. Recall that this scenario is very conservative (risk aversion to cancel cases is equal to 100%), therefore less cases are scheduled to prevent from canceling.

The total number of patients scheduled remains very similar regardless of the scenario when ICU availability is included. This number varies between 637 and 660 for instance 1 and between 597 and 619. Average wait times (μ_{wait}), standard deviation (σ_{wait}) (and therefore δ_{wait}) are again comparable.

Regarding the last metric (number of patients that may be canceled), Table 7 summarizes the results for both instances on all scenarios. For each instance, the first six lines represent the

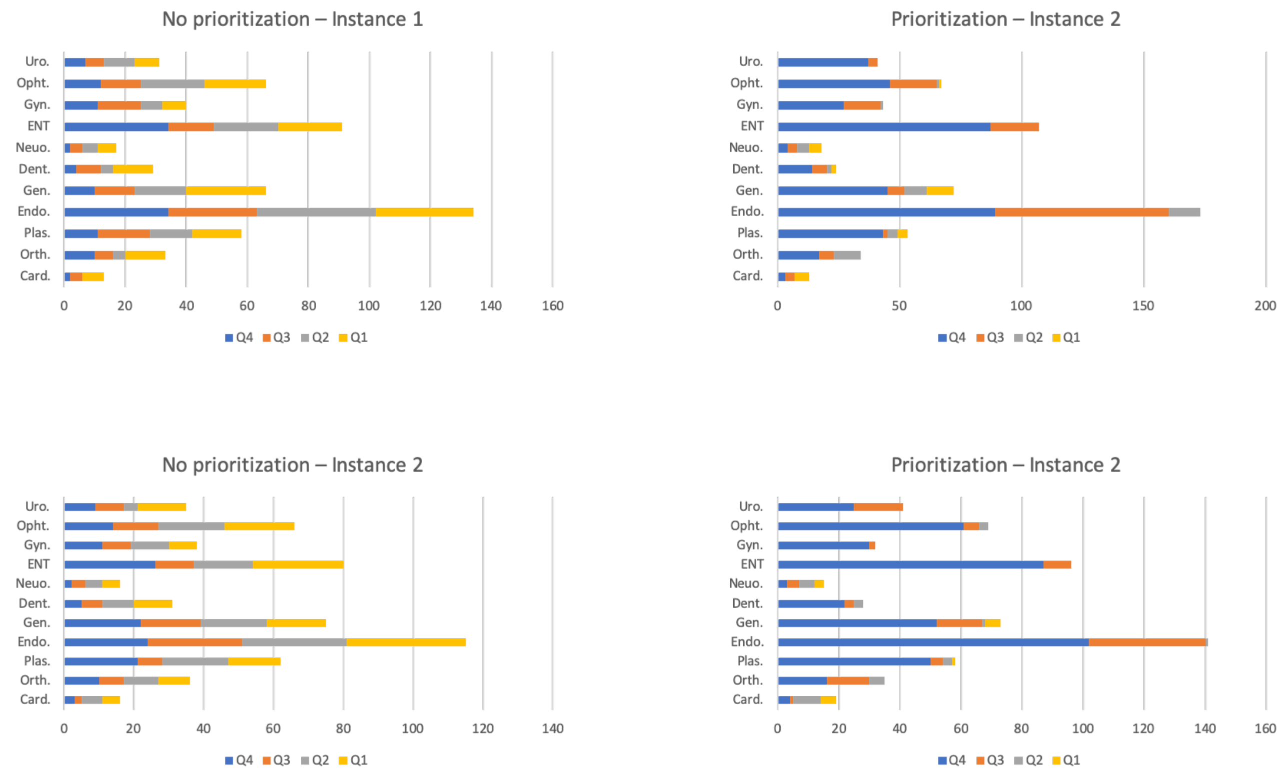


Figure 12: Distribution of number of patients in Q_1, Q_2, Q_3, Q_4 per specialty

number of elective cases scheduled requiring an ICU admission for different alternatives to include the availability in the ICU: with the static constraint or with the risk of cancellation. The next six lines detail the number of cases canceled after generating 5,000 simulations of probabilities to cancel cases.

Let's first note that the total number of patients programmed on the horizon is higher when the only constraint imposed on intensive care is a maximum of two patients daily. Next, our model is more likely to program patients when the probability to cancel patients is low (P_{low}). We observe that with the thresholds imposed ($u_1 = 0.2$ and $u_2 = 0.05$), our scheduling policy is more conservative and allows to schedule 2 patients requiring the ICU per day in very few occasions. In most cases, less patients are scheduled but for exceptional cases. Finally, imposing a constraint on intensive care has a smaller impact when the probabilities to cancel cases are low. The number of patients requiring the ICU per month varies from 0 (when the risk to cancel cases is very high) to 38 for the static constraint. The second half of the table shows that the number of patients to be canceled after 5,000 simulations increases as the probabilities used decrease: for example, being less conservative may lead to a higher cancellation of cases.

Table 7: Total demand of ICU beds (monthly)

		Instance 1			Instance 2		
	Scenario	P_{medium}	P_{low}	P_{high}	P_{medium}	P_{low}	P_{high}
Number of cases scheduled	Static	34	34	34	24	24	24
	$1-1$	21	21	21	38	38	38
	$0.5-0.25$	20	24	21	38	38	31
	$0.3-0.15$	18	25	16	31	32	19
	$0.2-0.05$	17	15	12	19	27	12
	$0-0$	0	4	0	0	4	0
# cancellations	Static	3.03	2.02	4.71	8.83	5.51	12.11
	$1-1$	2.67	2.60	4.94	8.77	5.59	11.70
	$0.5-0.25$	2.81	2.88	5.15	8.66	5.44	8.85
	$0.3-0.15$	2.72	2.64	2.84	6.42	4.71	3.51
	$0.2-0.05$	1.39	1.02	1.60	1.78	1.90	1.68
	$0-0$	0.0	0.0	0.0	0.0	0.0	0.0

We illustrate in Table 8 one example (Instance 2 with P_{medium}, Static vs $u_1 = 0.5, u_2 = 0.25$) where approximately the same number of patients are scheduled (20 vs 24). It shows the results for each day of the period (excluding week-ends). We observe that when considering the dynamic constraint, two patients are scheduled two days in a row on two occasions only (day 2 and day 24) versus six.

Table 8: Distribution of patients scheduled - Instance 2

Day	1	2	3	4	5	8	9	10	11	12	15	16	17	18	19	22	23	24	25	26	Total
Static	2	2	2	1	0	0	2	2	2	0	0	1	1	2	0	1	2	2	2	0	24
$u_1 = 0.5, u_2 = 0.25$	1	2	2	0	1	1	2	1	1	0	0	2	0	1	1	1	0	2	2	0	20
	6 vs 7					5 vs 6					4 vs 4					5 vs 7					

In conclusion, our approach allows to take into account the risk of cancellation more accurately than when we simply set a static constraint for each day. The number of patients scheduled on a day depends on the state of the system the previous day, which also depends on the day prior to that one, and so on. No scheme is therefore decided a priori for the week. In this way, the system is more flexible in the sense that there may be 2 per patient, regardless of the day of the week, but this value ultimately depends on the selected patients and the state of the system from the previous days. The constraint is thus transferred from the number of patients to operate per day to the probability of having cancellations.

6 Conclusion

In this paper, we have proposed a novel approach to model availability of beds in the ICU for elective surgery patients. This approach is inspired by Markov decision processes. We solve the surgical case assignment problem in the operating room at the tactical level. Four weeks in advance, we determine which patients will be operated on. We show that prioritizing patients during this process only increases the quality of the schedule without decreasing the occupancy rate of the OR. This mix of patients is also consistent with the master surgical plan agreed on by senior management. Finally, we include the risk of cancellations in the ICU during the planning of theses cases. The classical way is to include a constraint that limits the number of beds. However, this number varies under different conditions dynamically. This will prevent from canceling cases in the OR. We also better distribute the patients who need ICU admission throughout the week.

An extension to this work consists of increasing the number of states to evaluate (for example cancelling 3 cases) and including the length of stay in the ICU.

References

Agnetis, A., Coppi, A., Corsini, M., Dellino, G., Meloni, C., and Pranzo, M. (2012). Long term evaluation of operating theater planning policies. Operations Research for Health Care, 1(4):95–104.

Bowers, J. (2013). Balancing operating theatre and bed capacity in a cardiothoracic centre. Health care management science, 16(3):236–244.

Cardoen, B., Demeulemeester, E., and Beliën, J. (2010). Operating room planning and scheduling: A literature review. European journal of operational research, 201(3):921–932.

Chow, V. S., Puterman, M. L., Salehirad, N., Huang, W., and Atkins, D. (2011). Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. Production and Operations Management, 20(3):418–430.

Cook, D., Kozan, E., and Mchardy, S. (2004). Modelling of intensive care units and operating theatre in public hospitals.

Fei, H., Chu, C., Meskens, N., and Artiba, A. (2008). Solving surgical cases assignment problem by a branch-and-price approach. International Journal of Production Economics, 112(1):96–108.

Fügener, A., Edenharter, G. M., Kiefer, P., Mayr, U., Schiele, J., Steiner, F., Kolisch, R., and Blobner, M. (2016). Improving intensive care unit and ward utilization by adapting master surgery schedules. A&A Case Reports, 6(6):172–180.

Fügener, A., Hans, E. W., Kolisch, R., Kortbeek, N., and Vanberkel, P. T. (2014). Master surgery scheduling with consideration of multiple downstream units. European journal of operational research, 239(1):227–236.

Guerriero, F. and Guido, R. (2011). Operational research in the management of the operating theatre: a survey. Health care management science, 14(1):89–114.

Hulshof, P. J. H., Kortbeek, N., Boucherie, R. J., Hans, E. W., and Bakker, P. J. M. (2012). Taxonomic classification of planning decisions in health care: a structured review of the state of the art in or/ms. Health Systems, 1(2):129–175.

Leeftink, G. and Hans, E. (2018). Case mix classification and a benchmark set for surgery scheduling. Journal of scheduling, 21:17–33.

Macario, A. (2006). Are your hospital operating rooms "efficient"?: A scoring system with eight performance indicators. Anesthesiology, 105(1):237–240.

Marmor, Y. N., Rohleder, T. R., Huschka, T., Cook, D., Thompson, J., and Clinic, M. (2011). A simulation tool to support recovery bed planning for surgical patients. In Simulation Conference (WSC), Proceedings of the 2011 Winter, pages 1333–1339. IEEE.

Marques, I. and Captivo, E. (2017). Different stakeholders' perspectives for a surgical case assignment problem: Deterministic and robust approaches. European Journal of Operational Research, 261(1):260–278.

Mateus, C., Marques, I., and Captivo, E. (2018). Local search heuristics for a surgical case assignment problem. Operations Research for Health Care, 17:71–81.

McConnell, K., Richards, C., Daya, M., and Lowe, R. (2004). Effect of increased icu capacity on length of stay in the emergency department. Annals of Emergency Medicine, 44(4):S8.

Pham, D. and Klinkert, A. (2008). Surgical case scheduling as a generalized job shop scheduling problem. European Journal of Operational Research, 185:1011–1025.

Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., and Rademakers, F. E. (2016). Scheduling operating rooms: achievements, challenges and pitfalls. Journal of Scheduling, 19(5):493–525.

Santibanez, P., Begen, M., and Atkins, D. (2007). Surgical block scheduling in a system of hospitals: An application to resource and wait list management in a british columbia health authority. Health Care Management Science, 10:269–282.

van Oostrum, J., Van Houdenhoven, M., Hurink, J., Hans, E., Wullink, G., and Kazemier, G. (2008). A master surgery scheduling approach for cyclic scheduling in operating room departments. OR Spectrum, 30(2):355–374.

Vissers, J., Adan, I., and Bekkers, J. (2005). Patient mix optimization in tactical cardiothoracic surgery planning: A case study. IMA Journal of Management Mathematics, 16:281–304.

APPENDIX

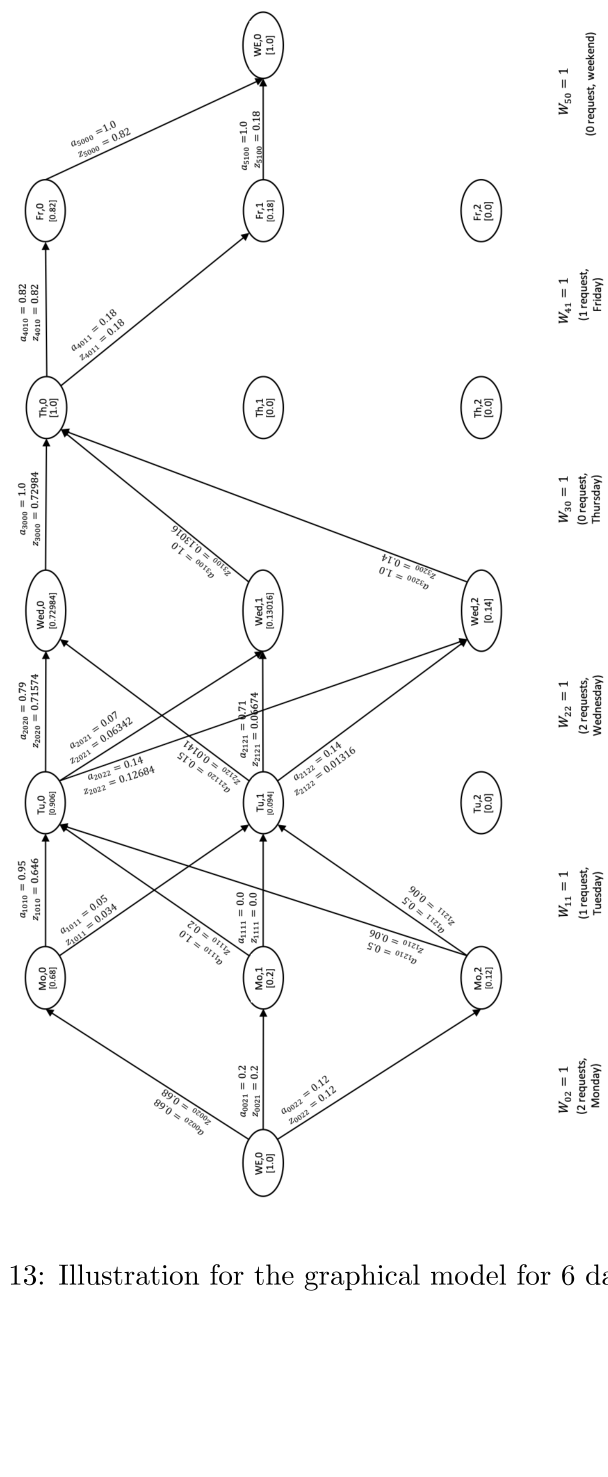


Figure 13: Illustration for the graphical model for 6 days