



Long-term integrated surgery room optimization and recovery ward planning, with a case study in the Brazilian National Institute of Traumatology and Orthopedics (INTO)



Cecília L. Siqueira^a, Edilson F. Arruda^{a,*}, Laura Bahiense^a, Germana L. Bahr^b,
Geraldo R. Motta^b

^a Industrial Engineering Program, Alberto Luiz Coimbra Institute - Graduate School and Research in Engineering, Federal University of Rio de Janeiro, Caixa Postal 68507, Rio de Janeiro, RJ 21941-972, Brazil

^b Instituto Nacional de Traumatologia e Ortopedia, Rio de Janeiro, Av. Brasil 500, São Cristovão, Rio de Janeiro RJ, 20940-070, Brazil

ARTICLE INFO

Article history:

Received 4 October 2015

Accepted 10 September 2016

Available online 15 September 2016

Keywords:

Healthcare modeling

Mathematical programming

Queuing theory

Scheduling

ABSTRACT

This paper proposes an integrated approach for the long-term planning and surgery allocation problem with downstream constraints. It is motivated by a case study in the Brazilian National Institute of Traumatology and Orthopedics, which provides elective high complexity surgeries for patients from the Brazilian public health system. We introduce an optimization problem that designs a periodic surgery allocation schedule as well as a recovery ward utilization plan, with a view at balancing patient arrivals and releases in the long term, in such a way that all surgeries are performed in a timely manner.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

Surgery allocation policies play a very important role in hospital management, for they can have a crucial, hard to predict impact in many of the services that are offered in a hospital (e.g., Belien & Demeulemeester, 2007). Inadequate policies can lead to mismatches between input and output rates in downstream processes, such as the recovery ward. That, in turn, can lead to an excess of surgery cancellations, thus compromising the system's capacity. Mismanagement can also lead to increased queues in the upstream processes, causing patients to wait too long for surgery.

There is a wide literature on surgery allocation problems. Most of the works, such as that of Min and Yih (2010a), deal with elective surgeries, which are scheduled in advance and thus facilitate the task of the decision maker. Other works, such as that of Bowers and Mould (2004), address non-elective and emergency surgeries, which typically require more intricate models. While most of the literature on surgery allocation problems employ mixed integer models (e.g., Agnetis et al., 2014), many other modeling tools are employed from the field of operations research (e.g., Hillier & Lieberman, 2014). Stochastic programming has been used

to address the probabilistic nature of variables such as the patient arrival and bed occupancy times (e.g., Min & Yih, 2010b), whereas other techniques can also be applied which exploit particular characteristics of the problems being studied, such as branch-and-price (e.g., Belien & Demeulemeester, 2008), heuristics (e.g., Belien & Demeulemeester, 2007) and simulation (e.g., Bowers & Mould, 2004). In another line of research, Pham and Klinkert (2008) applied multi-mode blocking job shop, a technique rather popular in industrial applications, to minimize the total time required to complete a set of prescribed surgeries. For a thorough treatment of surgery allocation problems, we refer to the excellent works of Cardoen, Demeulemeester, and Belien (2010), Hans and Vanberkel (2012) and Demeulemeester, Belien, Cardoen, and Samudra (2013).

The surgery scheduling and planning problem can be addressed in many different ways, depending on the objectives of the decision maker. One line of research focuses on the surgical center, striving to promote a better usage of it for some prescribed time window (e.g., Aringhieri, Landa, Soriano, Tanfani, & Testi, 2015a; Marques, Captivo, & Pato, 2012; Molina-Pariente, Fernandez-Viagas, & Framinan, 2015). Focusing on a single resource, however, often does not result in the best strategy for the whole system (e.g., Min & Yih, 2010b), and some authors choose a more detailed model of the system. A study by Blake (2010) showed that the bottleneck can actually be the number of available beds in the recovery ward, whereas Jonnalagadda, Hariharan, Walrond, and Prasad (2005) found out that the unavailability of recovery beds

* Corresponding author. Fax: +55 21 2270 9702.

E-mail addresses: cecilialorenzo@gmail.com (C.L. Siqueira), efarruda@po.coppe.ufrj.br, efarruda@gmail.com (E.F. Arruda), laura@po.coppe.ufrj.br (L. Bahiense), gbahr@into.saude.gov.br (G.L. Bahr), geraldomotta@terra.com.br (G.R. Motta).

accounted for 15% of the cancellations in a particular case study. Recovery beds occupation rates were evaluated in (e.g., Augusto, Xie, & Perdomo, 2010; Belien & Demeulemeester, 2007; Saadoui, Jerbi, Dammak, Masmoudi, & Bouaziz, 2015; Santibanez, Begen, & Atkins, 2007). Both Belien and Demeulemeester (2007) and Saadoui et al. (2015) studied problems with stochastic patient arrivals; Santibanez et al. (2007) studied a problem with multiple hospitals and Augusto et al. (2010) evaluated the effect of allowing patient recovery at the operating room if the recovery ward is full. Typically, a *closed block* paradigm is employed (e.g., Aringhieri et al., 2015a; Hans & Vanberkel, 2012; Aringhieri, Landa, & Tanfani, 2015b), whereby operating room (OR) sessions are assigned to specialties; in that case, each room session is assigned to a single surgery specialty. Sometimes, however, an *open block* paradigm may be used, whereby OR sessions can be comprised of multiple specialties, given that some rooms may be compatible with various surgery specialties. Not as popular as the closed block approach, the open block paradigm was investigated by Liu, Chu, and Wang (2011). In this paper, we strive to assess the gains provided by the open block paradigm in relation to its more popular closed block counterpart. The gains are evaluated in the light of some numerical examples, which suggest that they may be substantial in terms of effective room occupation rates, and cleaning and setup times.

Another critical issue arising in surgery allocation problems is the waiting queue for each specialty. Congested queues are often the major source of concern both for patients and hospital managers (e.g., Demeulemeester et al., 2013), for they can lead to long waiting times and the consequent deterioration of the patients health before surgery can take place. Bhattacharjee and Ray (2014) argue that a careful and structured patient flow management can reduce congestion and significantly improve hospital performance. A typical solution for queuing management is to assign priority to patients (e.g. Min & Yih, 2010a). Such a measure often leads to shorter waiting times for high priority patients. Other authors, like Aringhieri et al. (2015a) and Tanfani and Testi (2010), applied this approach by developing heuristics to select patients who should have priority in the surgeries planning. In addition to the prioritization of patients, Aringhieri et al. (2015b) also took into account the level of occupation of beds for planning surgeries, striving to promote a leveled occupation. However, all these approaches have failed to address a fundamental issue, which is ensuring that all patients are served in a timely manner. In this paper, we strive to propose an optimization model for the whole process, from patient arrival to patient discharge, taking into account the capacity constraints of the surgical center and of the recovery ward, with a view at optimizing the usage of the surgical center, while also keeping all queues within reasonable bounds in the long-term, and thus ensuring a timely service for all patients.

More specifically, this paper is motivated by a case study from the Brazilian National Institute of Traumatology and Orthopedics (INTO), which performs elective surgeries in a wide range of specialties. We propose a mathematical model for an elective surgery planning and allocation problem taking into account the entire process: from patient arrival to patient release from the recovery ward, after surgery is performed. Unlike previous studies, this paper addresses the long-term problem of balancing surgery offer and demand, in such a way that the waiting queues are kept within reasonable bounds. Such a long-term behavior is typically ignored in the literature, where short term performance measures, such as the hospital profit or productivity for prescribed planning periods, are often preferred. That, nonetheless, is often a myopic view, which may render the system unstable by favoring some surgical specialties at the expense of others, and thus producing increasingly higher waiting times for neglected specialties.

We accomplish long-term queuing management while also maximizing the utilization of the operating rooms for a prescribed

planning period (such as a week) by prescribing a master schedule to be repeatedly performed over the successive planning periods. We argue that maximizing the utilization of the operating rooms is beneficial in the long-term, since it results in a better utilization of the resources. However, it may be not sufficient to guarantee long-term stability, i.e. a sustainable balance between input and output patient flows. We achieve such a balance by means of constraints based on well-known queuing theory results, which are included in the optimization model to make sure every surgery is performed in a timely manner.

INTO's operating rooms can be used by distinct specialties, and that opens up the possibility of evaluating open block sessions. We formulated the model with open and closed block sessions, to provide insights into the effects of the latter on the performance, and found that significant gains can be obtained. To the best of our knowledge, this is a novel contribution of the present work.

Finally, considering that the queuing management problem is connected not only with the operating room occupation rates, but also with the bed management at the recovery ward, the model also generates a bed allocation plan. Such a plan divides the available beds among the surgical specialties, in such a way that the average number of patients released from the recovery ward in a given planning period surpasses the average number of patients entering the system over that same length of time. The rationale for this comes from well known queuing results, employed here to ensure long-term stability for each queue in the system.

To sum up, we stress that, even though motivated by a case study, the model is developed for a general elective surgery scheduling and planning problem. Hence, it offers a general framework for integrated surgery planning and scheduling, with multiple surgery specialties that are serviced in one or more hospitals.

This paper is organized as follows. Section 2 provides a brief explanation of INTO's operation. Section 3 introduces the mathematical model; Section 3.1 comprises some mathematical results intended to motivate the model for the whole system, which is introduced in Section 3.2. Section 4 is concerned with evaluating the long-term behavior of the system, should the solution be implemented. Finally, Section 5 features numerical simulations, and Section 6 concludes the paper.

2. Problem description

Located in Rio de Janeiro, the Brazilian National Institute of Traumatology and Orthopedics (INTO), briefly introduced in Section 1, is a public institution specialized in complex surgical care in both traumatology and orthopedics, that exclusively cares for patients from the Brazilian public health system. Although INTO does not provide first-aid services, it offers a trauma center that welcomes patients from other public health units. INTO covers a broad range of 13 orthopedic surgical specialties, called centers: *Pediatric, Feet and Ankle, Hand, Hip, Knee, Spine, Shoulder and Elbow, Microsurgery, Limb Lengthening, Oncology, Skull-Maxillo-Facial, Adult Trauma and Elderly Trauma*. Most of the eligible patients reside in the state of Rio de Janeiro, but the Institute also receives difficult cases from other Brazilian states; they belong to the Brazilian public health system and have their need for surgical treatment confirmed by a specialist physician, of one of the 13 centers.

Incoming patients are assigned an appointment with a physician from INTO's staff for a preliminary evaluation, which defines whether they are eligible to INTO's services. Eligible patients are then directed to one of the 13 existing centers, according to their condition. The allocated center then places the incoming patient in a queue, served in chronological order of arrival. After being serviced, the patient is sent to a recovery ward and stays there

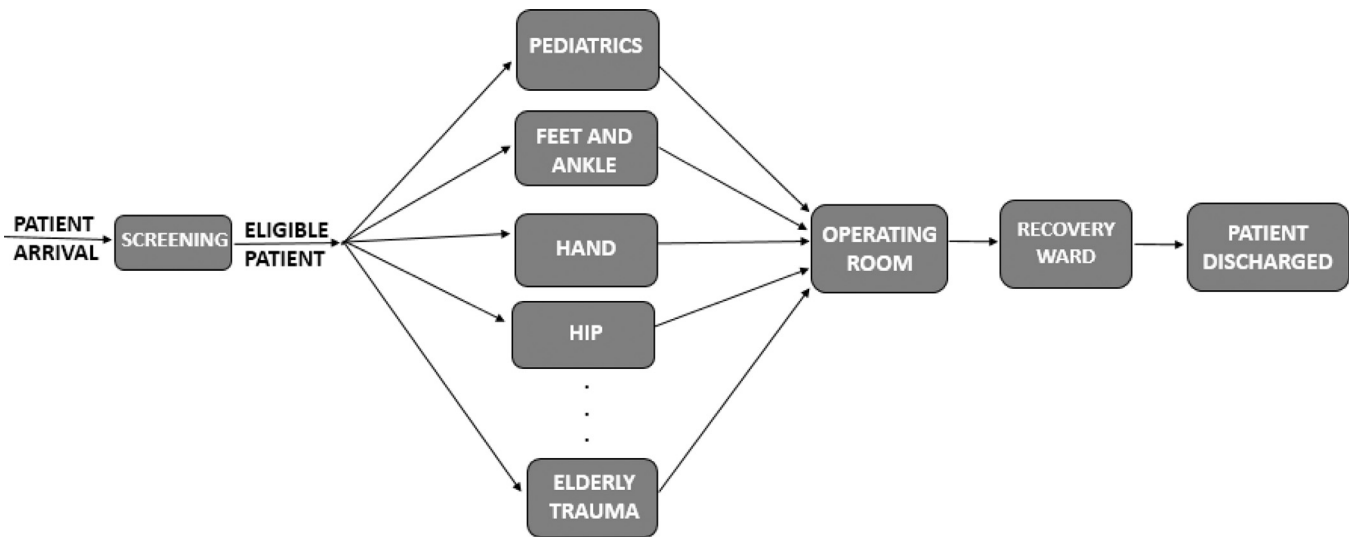


Fig. 1. INTO's patient flow.

until ready to be discharged. The whole process is illustrated in Fig. 1.

INTO's surgery rooms are distributed in two facilities. The main surgical unit has a total of 18 surgery rooms. Currently, 15 rooms are in activity and the others are used as support rooms. An auxiliary facility, called Day Hospital, holds a total of 3 operating rooms. Each of the 15 rooms in the main facility is capable of holding all available surgical procedures, whereas the surgery rooms at Day Hospital are equipped to serve only 4 surgical specialties, namely, Feet and Ankle, Hand, Limb Lengthening and Oncology.

Each surgery room opens for a total of 9 daily hours, from Monday to Friday. A setup time is needed to prepare the room for an incoming surgery, and a cleaning up time is necessary following each surgery. It is worth pointing out that each room can be set up in advance for its first surgery of the day. Moreover, the cleaning up following the last daily surgery can be performed in the after hours. Cleaning and setup times may vary from surgery to surgery but together they typically account for half an hour.

A recovery ward, comprised of 273 beds is available for after-surgery care. Of this total, 255 beds are located in the main facility and the remaining 18 beds are available at Day Hospital. Each bed in the main facility can serve patients from all surgery specialties, while beds in the Day Hospital can only serve the four specialties held there.

To put forth a strategic plan for surgery assignment, in such a way that the waiting queues are kept in acceptable levels, is not a trivial task. Indeed, INTO has struggled with ever increasing queues that resulted from the former policy of maximizing the total number of surgeries, paying no heed to the balance between arrivals and departures. An integrated approach, detailed in Section 3, is applied to INTO's problem, with a view at tackling their mismatch between arrivals and departures.

A novel feature of the model is that, besides ensuring a desirable long-term behavior by tackling the queuing problem resulting from the input/output processes of each queue, it also addresses the downstream process resulting from sending the patients to the recovery wards. That is necessary to ensure the feasibility of the surgery schedules, since no patient can be operated unless there is a place available for them in the recovery ward at the end of their surgery. The objective is to provide the decision maker with a surgery plan that ensures long-term stability, taking into account the limited capacity of both the surgery rooms and the recovery ward.

3. Mathematical formulation

This section starts with some preliminary remarks regarding the problem and the model, to motivate the model choices introduced in Section 3.2.

3.1. Preliminaries on the formulation

Consider the problem introduced in Section 2. It is not difficult to see that, for each surgical specialty, the patient flow in the system, from their admission into the waiting list to their release from the recovery ward, can be modeled as a queuing system. The performance of such a system depends on the long term relationship between the admission and discharge rates which, ultimately, determine the occupation rate of the queue, defined as:

$$\rho_s = 1 - \lim_{t \rightarrow \infty} P(X_s(t) = 0), \quad s \in S, \quad (1)$$

where $S = \{1, \dots, N_s\}$ is the set of surgical specialties; $\{X_s(t), t \geq 0\}$ is a stochastic process representing the number of patients in queue s as time elapses; and $P(X_s(t) = 0)$ stands for the probability that queue s is empty at time t . Hence, the occupation rate of the queue represents the probability that there are patients waiting for surgery. In particular, $X_s(t) = w$ indicates that w patients are either being serviced or waiting for service at time t in queue s . It is well known that, as the value of ρ_s approaches the unit, the system becomes increasingly congested and the queues increase exponentially (e.g., Kleinrock, 1975; Hillier & Lieberman, 2014).

Providing that our modeling efforts are more meaningful when the systems' performance is unsatisfactory, we assume, for the sake of argument, that the values of $\rho_s, s \in S$, are arbitrarily close to one, in such a way that:

$$P(X_s(t) < m) \approx 0, \quad \text{for some } m < \infty, \quad \forall t > 0, \quad s \in S. \quad (2)$$

Let us suppose that m is large enough such that, for each specialty, there are enough patients in the waiting list to carry on any feasible periodic surgery schedule. We assume that $m < \infty$ to ensure that the system is stable, i.e. that the queues are kept finite at all times.

Remark 1. Define an aggregate stochastic process $\{X_t, t \geq 0\}$, to represent the total expected number of surgery hours for all the patients waiting in line for all surgical specialties. In particular, $X(t) = n$ indicates that there are n surgery hours in the queue at time t . Then:

- (a) If the recovery ward is large enough to receive patients coming from surgery at any time, the optimal strategy with respect to minimizing $\{X_t\}, t \geq 0$, is to apply, at all periods, the feasible schedule that maximizes the number of surgery hours.
- (b) If the capacity of the surgery rooms is more than enough to match the demand rates of all queues, the optimal plan with respect to $\{X_t\}, t \geq 0$, is the best plan to share the ward's capacity among all the surgical specialties, such that the output rate of patients at the recovery ward surpasses the input rate for all available specialties.

The first alternative of Remark 1 refers to the situation when the congestion leading to Eq. (2) is caused by the lack of capacity in the surgery rooms, or by the mismanagement of the available capacity. The second alternative happens when the congestion that leads to Eq. (2) is caused by the lack of capacity in the recovery ward, or by the mismanagement of the ward's available capacity.

The model in the next section takes into account both scenarios, and produces a periodic surgery schedule aimed at tackling the first problem, along with a recovery ward bed allocation plan aimed at tackling the second one. Even if there is shortage of neither surgery room capacity nor recovery ward beds, the model in Section 3.2 promotes a better long term management of the capacity, such that the decision maker can tackle more efficiently possible increases in the demand rate for surgeries.

Note that the model is deterministic, aimed at solving the long-term management of input and output flow of patients. Such a formulation ensures long-term stability, i.e. makes sure that all queues are kept within finite bounds in the long term, by ensuring that the output rate of patients surpasses the respective input rate for all specialties. However, the actual performance of the queues will depend on factors such as the spare capacity in the surgery rooms and in the recovery ward, as well as on the parameters of the stochastic process describing the demand for each specialty and the respective recovery time for patients belonging to that specialty.

3.2. The model

To model the overall process, from the service queues to the recovery ward, we define a set $I_1 = \{1, \dots, N_{i1}\}$ of available operating room (OR) blocks at INTO; each block is characterized by a pair of indices, denoting the OR and the day. Similarly, we define a set $I_2 = \{1, \dots, N_{i2}\}$ of OR blocks at Day Hospital, and let $I = I_1 \cup I_2$. Recall that $S = \{1, \dots, N_s\}$ was defined in Section 3.1 as the set of surgical specialties. Define also a set $J = \{1, \dots, N_j\}$ of candidate surgeries for a given planning period. Typically, the set J is comprised of a prescribed number $n_s > 0$ of surgeries of each specialty s , in such a way that

$$\sum_{s \in S} n_s = N_j.$$

Observe that any periodic plan of d days can be viewed, without loss of generality, as an equivalent daily plan for a hospital d times as large, with the same number of OR blocks. The solution can be mapped back into a periodic plan by recovering the OR and day indices of each block in the original periodic formulation.

The model has four distinct groups of decision variables. The first group comprises the binary variables $x_{i,j}$, that are assigned the value 1 whenever surgery $j \in J$ takes place in room $i \in I$ and are set to zero otherwise. The second group contains the binary variables $L_{i,j,r}$, which receive value 1 whenever surgeries j and r are both assigned to room i and surgery j precedes surgery r , and are set to zero otherwise. The third group is comprised of continuous positive variables S_j , which represent the starting time of surgery j , for all $j \in J$. The last decision variables belong to the set $Y_{s,s} \in$

S , which define an integer number of beds to be assigned to specialty s in the recovery ward. Cleanup and setup times may differ depending on the surgery types; hence, under the open block approach the order of the surgeries determines the feasibility of the schedules, and not all permutations of a group of surgeries will be feasible. That motivates the inclusion of the second and third groups of decision variables in the model.

According to the analysis in Section 3.1, the objective is to define a daily surgery schedule in such a way that the actual number of surgery hours is maximized. Note that this is equivalent to minimizing the idle time over the whole set of OR blocks. Thus, the objective function can be defined as:

$$\text{Max} \sum_{i \in I} \sum_{j \in J} \text{Time}_j \cdot x_{i,j}, \quad (3)$$

where the parameters Time_j are obtained from an input table containing the expected surgery times for all surgeries $j \in J$.

The first constraint of the model establishes that each surgery is executed at most once. That is equivalent to saying that each surgery can be allocated to at most one OR block, thus resulting in the following set of inequalities:

$$\sum_{i \in I} x_{i,j} \leq 1, \quad \forall j \in J. \quad (4)$$

The relation between decision variables $x_{i,j}$ and $L_{i,j,r}$ is established in inequalities (5) and (6).

$$L_{i,j,r} + L_{i,r,j} \leq \frac{1}{2} * (x_{i,j} + x_{i,r}), \quad \forall i, j, r \mid j \leq r \quad (5)$$

$$L_{i,j,r} + L_{i,r,j} \geq x_{i,j} + x_{i,r} - 1, \quad \forall i, j, r \mid j \leq r \quad (6)$$

Inequalities (5) and (6) act together, with j and r representing distinct surgeries belonging to set J . If none of these surgeries is allocated to room i , the right hand side of inequality (5) is set to zero, while the right hand side of inequality (6) equals -1 . Hence, since these are binary variables, the sum must be zero, which also means that both variables should be set to zero. If, on the other hand, only one of the surgeries j or r is allocated to room i , the right hand side of inequality (5) adds up to 0.5, whereas the right hand side of inequality (6) is nil, implying that both binary variables are nil. As a consequence, these surgeries are not allocated to the same room. Now, if both surgeries are allocated to room i , the right hand sides of both inequalities (5) and (6) add up to one, implying that only one of these binary variables is set to one. Surgery j precedes surgery r in room i if $L_{i,j,r} = 1$ and the opposite is true whenever $L_{i,r,j} = 1$.

Now, let us consider the following pair of inequalities, which are concerned with the starting times of the allocated surgeries:

$$S_r \geq S_j + \text{Time}_j + \text{Clean}_j + \text{Setup}_r - (H + \text{Clean}_j) \cdot (1 - L_{i,j,r}) \quad \forall i, j, r : j \neq r, \quad (7)$$

$$S_j \leq H - \text{Time}_j, \quad (8)$$

where H is total daily period available for surgeries at each surgery room. Inequality (8) only allows surgeries to start which can be finished within the daily window of H active hours for each room.

To interpret inequality (7), recall that $S_r, r \in J$, is the starting time of surgery r . Assume that surgeries j and r are allocated to the same room $i \in I$. Suppose, without loss of generality, that j is to start before r . Then, from inequalities (5) and (6), it follows that $L_{i,j,r} = 1$. Consequently, the starting time for surgery r , S_r , must be larger than S_j plus the execution time of surgery j , denoted as Time_j , plus the sum of the cleaning up time for surgery j (Clean_j) and the setup time for surgery r (Setup_r). If none of these surgeries is allocated to room i , constraint (7) will be deactivated for this

pair of surgeries, for the right hand side of this equation will become negative and hence S_r will be able to take any non-negative value.

An arbitrary number of patients from each queue can be added to the set J of potential surgeries to be performed. Typically, one can add as many surgeries of a single type as could be held in all the OR blocks together, had blocks been assigned exclusively for that type of surgery. The constraints in (9) are known as nonanticipativity constraints, and they are usually employed in a stochastic context, acting as coupling constraints and ensuring that the future cannot be anticipated. In our context, they have been used as symmetry breaking constraints, added to enforce a logical order in the execution of the same type surgeries. Saying otherwise, constraint (9) ensures that the $(n+1)$ th surgery of type j can only be initiated after the n th surgery of that type has begun. That is, one can only allocate surgery $n+1$ of type j to a room i if surgery n of that type has been already assigned to some room. These constraints are summarized in inequality (9).

$$\sum_{i \in I} x_{i, j_{s(n+1)}} \leq \sum_{i \in I} x_{i, j_{sn}} \quad \forall j_s, s \in S. \quad (9)$$

where $J_s \subset J$ is an ordered set of all the surgeries of type $s \in S$ contained in the candidate surgery set J , and j_{sn} is the n th element of that set.

As pointed out earlier, this paper also aims at addressing the problem of managing the long term queues for each specialty. To keep the queues from growing indefinitely, we establish that the minimum number of surgeries to be performed over a period for each type of procedure $s \in S$ exceeds the average input rate of patients for that procedure. This constraint, summarized in inequality (10), also aims at obtaining a balanced surgery schedule, which permits the reduction of the queues for all surgical centers:

$$\sum_{j \in J_s} \sum_{i \in I} x_{i,j} \geq \text{InputRate}_s, \quad \forall s \in S, \quad (10)$$

where $\text{InputRate}_s, s \in S$, are parameters provided to the decision maker and represent the average demand over a period for each specialty.

So far, the model addressed issues related to the demand for services and the capacity of the OR blocks. However, for a schedule to be feasible, one needs enough available beds in the recovery wards to send all the patients subject to surgery at any given period. To help the decision maker manage the bed availability in the recovery ward, we introduce a set of variables Y_s to represent the number of beds available in the recovery ward for each specialty $s \in S$. The management of the flow of patients from the surgery rooms to the recovery ward is represented in inequality (11).

$$\sum_{j \in J_s} \sum_{i \in I} x_{i,j} \leq \frac{1}{R_{Ts}} Y_s, \quad \forall s \in S, \quad (11)$$

where R_{Ts} is the average recovery time for patients of specialty $s \in S$, in periods. Hence, $\frac{1}{R_{Ts}}$ is the average number of patients release per period, per recovery bed, for specialty s . Consequently, the right hand side of the above inequality represents the total number of beds released periodically for specialty s , in average. As a result, inequality (11) makes sure that the number of allocated surgeries for any specialty does not exceed the number of beds released in the recovery ward for the same specialty at any period.

Note that constraints (10) and (11) are stated in terms of average requirements, which is consistent with the long-term analysis proposed in the paper, relying on well-known queuing theory results to ensure a long-term balance between input and output patient flows. Naturally, fluctuations will occur, but detailed operational-level analysis is beyond the scope of the paper.

Let N_{beds} denote the total number of beds in the recovery ward. Inequality (12) couples the decision variables representing

the number of recovery beds for each specialty, making sure that their sum does not exceed the total number of beds in the recovery ward.

$$\sum_{s \in S} Y_s \leq N_{\text{beds}} \quad (12)$$

Finally, there is a peculiarity with regards to the set of surgeries, since some surgeries can only be performed at INTO's main facility. Let $J_n \in J$ denote the subset of such surgeries. Equation (13) prevents any surgery from this set from being allocated to Day Hospital:

$$\sum_{i_2 \in I_2} \sum_{j \in J_n} x_{i,j} = 0. \quad (13)$$

We could have defined variables $x_{i,j}$ only for feasible allocations, but we chose the formulation above to make the model more versatile. If the peculiarity that some surgeries can only be performed at INTO ceases to exist, it would be enough to eliminate this type of constraints from the model, without having to change the rest of its structure.

The complete problem, which is formulated under the *open block* paradigm, can then be written as:

$$\begin{aligned} & \text{Maximize} \quad \sum_{i \in I} \sum_{j \in J} \text{Time}_j \cdot x_{i,j} \\ & \text{Subject to} \quad (4) - (13), \text{ and} \\ & x_{i,j} \in \{0, 1\}, \quad \forall i, j, \quad L_{i,j,r} \in \{0, 1\}, \quad \forall i, j, r, \\ & S_j \in \mathbb{R}^+ \quad \forall j, \quad Y_s \in \mathbb{Z}^+ \quad \forall s. \end{aligned} \quad (14)$$

The *closed block* formulation for the same problem needs an additional constraint, namely:

$$L_{i,j,r} + L_{i,j,r} = 0, \quad \forall i \in I, j \in J_{s_1}, r \in J_{s_2} : \{j \neq r, \text{ and } s_1 \neq s_2\}, \quad (15)$$

where $s_1, s_2 \in S$ are surgical specialties and $J_s \in J$ is the subset of all candidate surgeries of specialty $s \in S$. Constraint (15) does not allow different surgery types in the same OR block.

It is worth pointing out that the approach, while developed for two hospital facilities, can be easily generalized for multi-facility problems. That can be accomplished by adding incompatibility constraints in the form of (13) for all available facilities.

Remark 2. Please note that, as a modeling choice, the surgery starting times are outputs of the model. Such a modeling choice helps us seamlessly deal with potential situations where there may be significant mismatches between setup and/or cleaning times in the open block approach. The constraints regarding the surgery times in (11), however, are stated in terms of average requirements. Whereas ensuring long term stability, these constraints are not concerned with the operational level. If one wishes to provide further protection against peak bed occupancies, one can tweak the parameters in (11) – for instance replacing the average recovery times with appropriate quantiles. These tweaks can be evaluated by means of a detailed queuing analysis (see Section 5.3), until the resulting queuing behavior is satisfactory for all specialties.

4. Long term performance evaluation

For each specialty $s \in S$, the performance of the associated queue clearly depends on demand rate – InputRate_s – and on the output flow, which is determined by both the number of surgeries scheduled and the number of available beds for this specialty in the recovery ward. Indeed, if there is spare capacity in the surgery rooms, the release of patients from the queue will be limited by the number of beds available in the surgery ward which, as the bottleneck, will ultimately determine the output rate of patients from the queue. In such a case, the decision maker can employ the empty slots in the surgery rooms to add robustness to the model

and compensate for fluctuations in the output flow of patients from the recovery ward. As a result, one can model the queue for each specialty $s \in S$ as a $\text{---}/\text{---}/Y_s$ queue (e.g., Hillier & Lieberman, 2014), recalling that Y_s is the number of beds available for specialty s in the recovery ward. The blank spaces in the notation above will be filled depending on the distribution of both the time between successive arrivals to the queue and the time between successive releases from the recovery ward per bed, respectively. For example, if both the time between arrivals and the recovery time per bed were exponentially distributed, the system would be described by an $M/M/Y_s$ queue.

If, on the other hand, there is spare room in the recovery ward, it is the surgery rooms that will limit the output flow of patients. In such a case, the decision maker can take advantage of the slack capacity in the recovery ward to compensate for fluctuations in the length of stay of patients in the recovery ward. In that case, the performance of the queuing system related to some specialty $s \in S$ depends on its input rate – InputRate_s – and on the number of surgeries available for this specialty. As a result, one can model the queue for each specialty $s \in S$ as a $\text{---}/D/dp(s)$, where $dp(s)$ is the number of surgeries for specialty $s \in S$ over a period. The blank space in the notation will be filled depending on the distribution of the times between patient arrivals for each specialty, and the service is assumed to be deterministic because the daily plan prescribes a deterministic number of surgeries of a given specialty.

5. Numerical experiments

In this section, we employ the model to INTO's case, both in open and closed block form; the comparison of results is presented in Section 5.1. Section 5.2 is concerned with the application of these outcomes at INTO, while Section 5.3 evaluates the long-term performance of the open block solution. Finally, Section 5.4 features a set of randomly generated examples, designed to provide a contrast between open and closed block solutions. For all experiments in Sections 5–5.2 and 5.4, we added patients from all specialties to the set J of potential surgeries to be performed. For each specialty, we added as many surgeries of this specialty as could be held in all the OR blocks together, had all blocks been assigned exclusively for that specialty.

We start by applying the model in Section 3 to INTO's surgery allocation and planning problem, briefly described in Section 2, with a view at comparing open and closed block formulations. Data concerning the average daily demand for each of their 13 specialties $s \in S$, denoted by InputRate_s , as well as the average surgery time for each specialty, which determines the surgery times for each candidate surgery $j \in J$, were provided. The hospital also provided the average release time from the recovery ward R_{Ts} for patients that underwent surgeries of each specialty $s \in S$, along with the number of surgery rooms, which determines the composition of the set I for a given planning period, as explained in Section 3.

For the example, we have $\alpha = 18$ surgery rooms available per day and a daily working period of $H = 9$ hours. Furthermore, INTO's physicians agreed on a total time of 30 minutes for cleaning up after a surgery and setting up for the next one, regardless of the surgeries taking place. Finally, the total number of beds currently available in their recovery ward $N_{\text{beds}} = 273$. The problem was implemented in AIMMS® 4.3.2 and solved using CPLEX® 12.6. We ran the code in an Intel-core i5 notebook with 2.3 gigahertz and 6 megabytes RAM, running Windows 7. The optimal solution was found in approximately 710 minutes.

We obtained the demand rate for each specialty ($\text{InputRate}_s, s \in S$) from a database containing the number of daily admissions per specialty from 2012 to 2014. To obtain an equivalent input rate per business day, the total number of patients over this period was divided by the total number of business days of this period.

Table 1
Input rate, surgery time and recovery time for INTO's centers.

Surgical center	Daily demand rate	Surgery time (hours)	Recovery time (days)
Pediatric	1.96	2.07	4.12
Feet and Ankle	2.94	2.11	4.11
Hand	5.98	1.79	2.22
Hip	4.24	2.88	11.18
Knee	6.47	2.92	6.28
Spine	2.63	4.57	13.05
Shoulder and Elbow	2.41	3.21	3.20
Microsurgery	0.99	2.28	6.00
Limb Lengthening	0.75	1.79	6.39
Oncology	1.29	1.81	4.29
Skull–Maxillo–Facial	1.34	3.16	3.15
Adult Trauma	2.02	2.79	9.29
Elderly Trauma	0.55	3.06	11.86

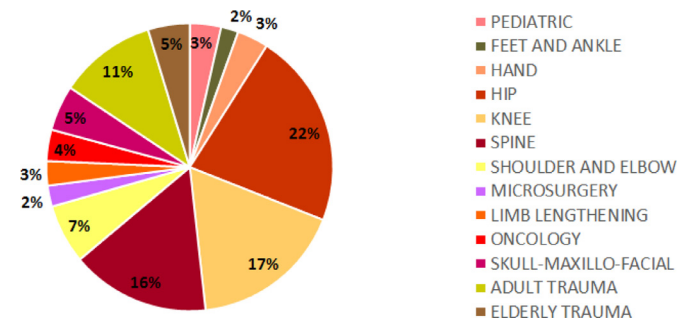


Fig. 2. Number of beds per surgical specialty for INTO.

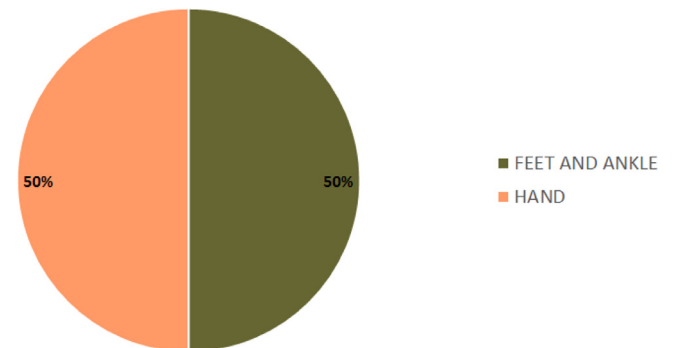


Fig. 3. Number of beds per surgical specialty for Day Hospital.

To compute the average surgery time per specialty we made use of data on the actual surgery time of each surgery of each specialty $s \in S$ performed from 2012 to 2014. The average time spent in the recovery ward for each specialty was also obtained from data corresponding to the year of 2014. The patients can be released in any day of the week, including the weekends, and the time they spend in the recovery ward does not discount the weekend days. However, since we are planning for the business days only, the recovery period R_{Ts} was multiplied by a factor of $\frac{5}{7}$ to be scaled into the average number of business days spent in the recovery ward per specialty.

Table 1 presents, in compact form, INTO's 13 surgery specialties and their respective parameters. Note that the demand rate is measured in patient arrivals per day, for each specialty.

We solved problem (14) to obtain an optimal daily surgery plan for INTO and Day Hospital, under the open block formulation. Fig. 2 summarizes the results concerning the optimal number of recovery beds for INTO obtained from the model for each specialty. Fig. 3 depicts the optimal number of beds for Day Hospital. Considering both hospitals, observe that only four specialties, namely

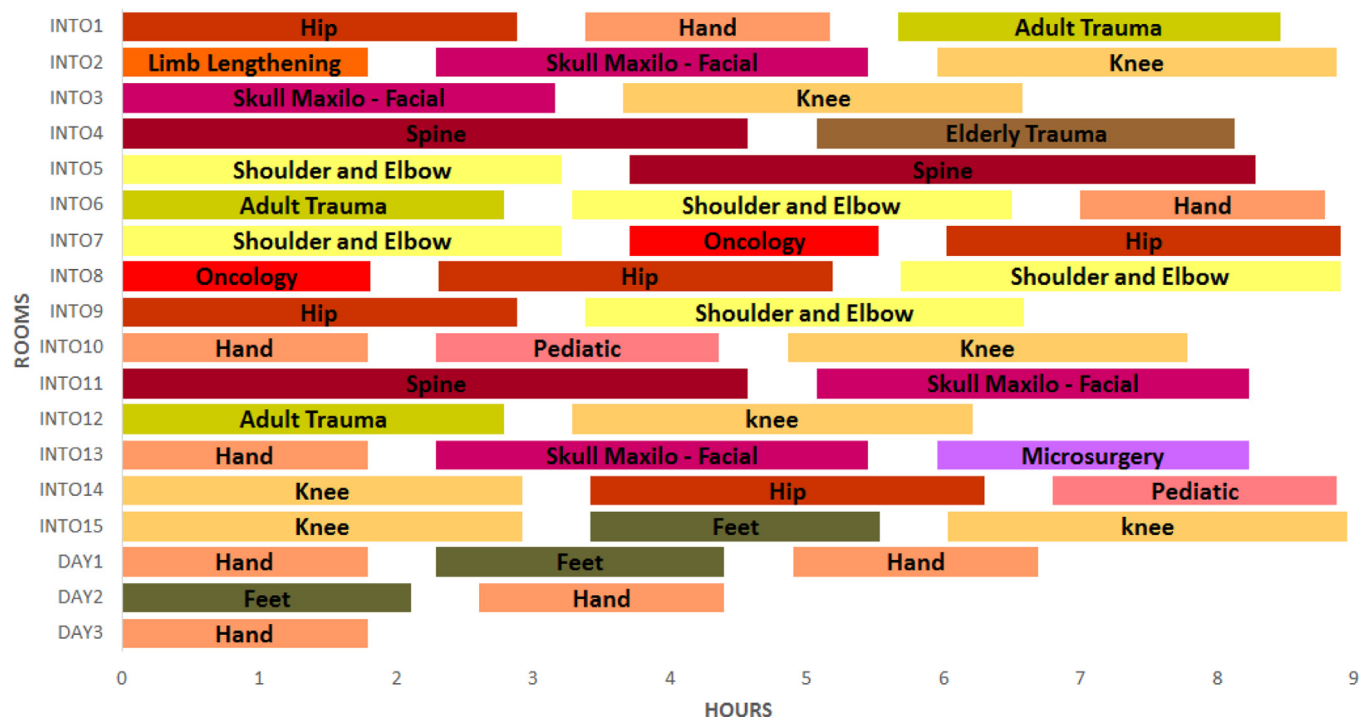


Fig. 4. Optimal daily surgery schedule for INTO.

Table 2
Daily surgeries and recovery beds per specialty.

Surgical center	Number of daily surgeries	Number of recovery beds
Pediatric	2	9
Feet and Ankle	3	14
Hand	7	18
Hip	5	56
Knee	7	44
Spine	3	40
Shoulder and Elbow	5	17
Microsurgery	1	6
Limb Lengthening	1	7
Oncology	2	9
Skull–Maxillo–Facial	4	13
Adult Trauma	3	28
Elderly Trauma	1	12

Hip, Knee, Shoulder and Elbow and Adult Trauma, account for 60% of the available beds.

The model also outputs the number of daily surgeries, as well as their disposition in the surgery rooms. Fig. 4 summarizes these results. Note that some rooms are left with considerable spare capacity. That is consistent with the fact that the bottleneck for our example is the number of beds in the recovery ward. Increasing the number of beds available for the after-surgery recovery, one can expect an increase in the occupation of the surgery rooms.

To summarize the open block results, Table 2 features the number of surgeries and the number of beds in the recovery ward per specialty.

5.1. Model analysis: closed block results and comparisons

We employed problem (14), with additional constraint (15), to obtain an optimal daily surgery plan for INTO and Day Hospital, under the closed block formulation. It turns out that the problem is infeasible. The minimum number of rooms to make the problem feasible is shown in Table 3, which also features the minimum number of rooms per specialty.

Table 3
Minimum number of rooms per specialty for feasibility.

Surgical center	Surgery time (hours)	Surgeries per room	Daily surgeries	Required rooms
Pediatric	2.57	3	3	1
Feet and Ankle	2.61	3	3	1
Hand	2.29	3	6	2
Hip	3.38	2	5	3
Knee	3.42	2	7	4
Spine	5.07	1	3	3
Shoulder and Elbow	3.71	2	3	2
Microsurgery	2.78	3	1	1
Limb Lengthening	2.29	3	1	1
Oncology	2.31	3	2	1
Skull–Maxillo–Facial	3.66	2	2	1
Adult Trauma	3.29	2	3	2
Elderly Trauma	3.56	2	1	1
Total				23

Now assume that INTO could expand to 23 surgery rooms, with 20 in the main facility and 3 at Day Hospital. Under this scenario, the solution is depicted in Fig. 5.

Comparing Figs. 4 and 5, one can see that the open block approach makes a better use of the available capacity for the proposed example. Indeed, a closed block model would require a 27.7% increase in capacity, and INTO would have to expand its surgical center to 23 rooms simply to attain feasibility. Moreover, while the open block model features a daily occupation rate of 73.68%, its closed block counterpart presents a 50.64% occupation rate. In addition, the open block model features a total of 44 daily surgeries, performed in 119.36 surgery hours, while the competing approach manages only 39 surgeries, performed within a time span of 104.84 hours. These results illustrate the increased efficiency of the open block approach with respect to the closed block approach.



Fig. 5. Closed block alternative daily surgery schedule for INTO.

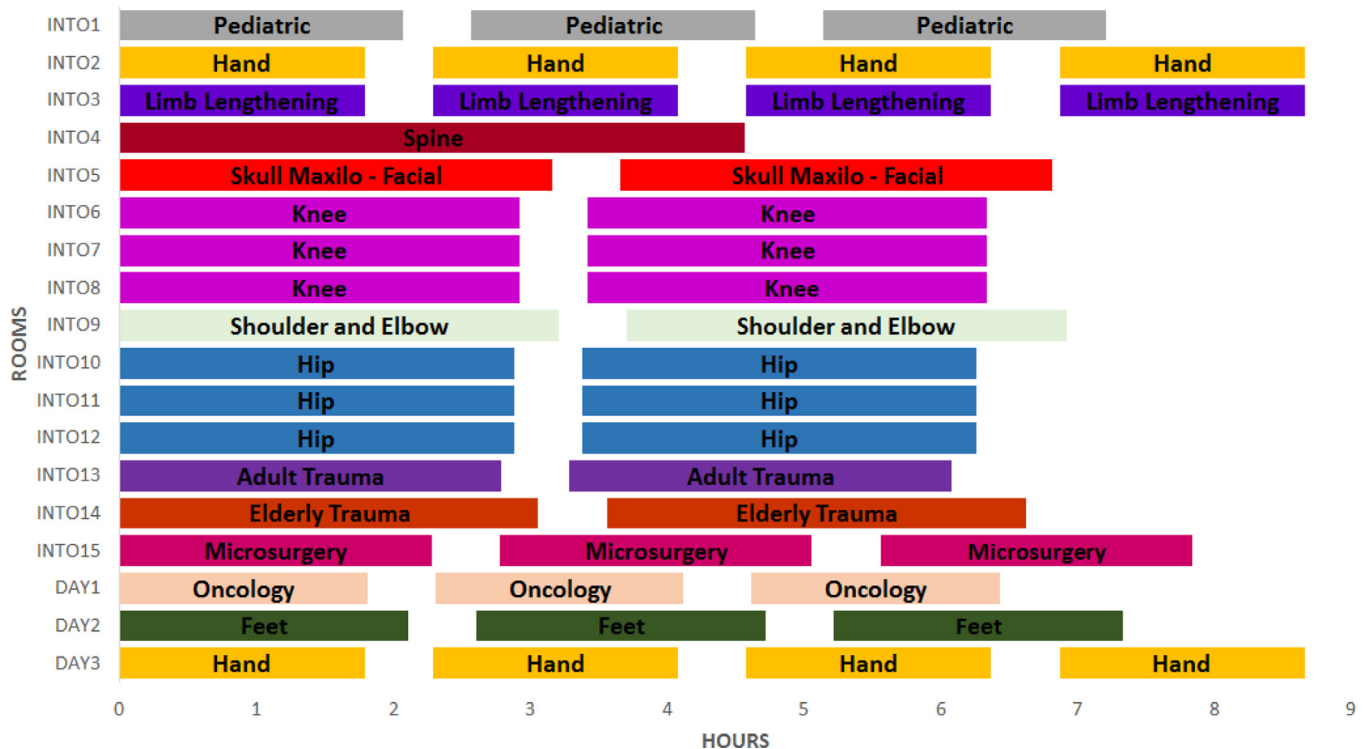


Fig. 6. INTO's policies.

5.2. Performance comparison: application of the approach at INTO

Before our study was undertaken, INTO's policy was to maximize the number of daily surgeries under a closed block approach, paying no heed to the actual demand for each specialty. The resulting schedule, depicted in Fig. 6, contemplates 38 surgeries and a total of 88.71 hours of surgery. Considering that the 18 rooms in

use yield a total of 162 working hours, that results in a 55% occupation rate.

After a careful inspection of Table 1, one can see that this schedule does not meet the demand of the following specialties: *Hip*, *Knee*, *Spine* and *Shoulder and Elbow*, which are specialties that typically require longer surgery and recovery times. Such a schedule resulted in ever increasing queues for these specialties.

Table 4
Number of daily surgeries at INTO before and after the model.

Surgical center	Surgeries INTO (before)	Surgeries INTO (after)	Surgeries proposed
Pediatric	3	2	2
Feet and Ankle	3	3	3
Hand	8	8	8
Hip	2	5	5
Knee	6	6	6
Spine	1	2	3
Shoulder and Elbow	2	3	5
Microsurgery	3	1	1
Limb Lengthening	4	1	1
Oncology	2	2	2
Skull–Maxillo–Facial	2	2	3
Adult and Elderly Trauma	4	5	4

Considering that the full implementation of the approach would require a thorough change of culture, INTO opted for a partial implementation of it. Table 4 summarizes the number of daily surgeries at INTO before and after the partial embrace of the approach. One can see that most specialties have more surgeries than incoming patients. However, *Knee* and *Hip* remain with slightly more incoming patients than surgeries.

The partial implementation of the approach at INTO had positive results in the first semester of 2015. For the first time in years, all queues were decreased, with more modest gains for the most complex surgeries. Unfortunately, there was a nurse strike during most of the second semester of 2015, resulting in increased queues for the overall year of 2015. INTO is currently considering a full implementation of the approach, either in open or closed block form.

5.3. Long-term queuing analysis

The model enables the decision maker to produce a detailed analysis of the queuing systems in the surgery management process. For the example, the system results in one queue for each specialty, whose characteristics depend on both the patient arrival process and the stochastic process underlying the bed occupation times for each specialty. Given that INTO serves a large population, that of the state of Rio de Janeiro – Brazil, and considering that the data was not stored in such a way that we can retrieve the distribution of the time between incoming patients or even the daily distribution of demand (the data are accumulated for arbitrary amounts of time and then entered to the system), we assume that the patients come according to a Poisson distribution, whose daily rate is retrieved from Table 1. Such a choice is motivated by the fact that often data for inter-arrival times collected from large populations tend to be exponentially distributed, generating a Poisson arrival process (e.g., Milne & Whitty, 1995; Pearson, Reynolds, & Stickley, 2012).

From the open block results in Section 5, one can see that all recovery ward beds are used, while some operating rooms are left with considerable spare capacity. Hence, considering the discussion in Section 4, we can assume that the service rate is the rate at which recovery ward beds are made available for incoming patients, who have just been subject to surgery. The data concerning the recovery period per specialty is also not available in such a way as to allow us to construct a probability distribution. Hence, for the sake of analysis, we also follow previous results in the literature (e.g., Milne & Whitty, 1995; Pearson et al., 2012), and assume that the number of patients released daily from the recovery ward follows a Poisson process, whose rate is recovered from the data as $\mu_s = \frac{1}{RT_s}$, $s \in S$. As a result, we have the elements to build a M/M/ Y_s queuing model for each specialty, from which we can recover performance measures such as waiting time for surgery and length of the queue in the long term. For each specialty $s \in$

S , the occupation rate is given by $\rho_s = \frac{\lambda_s}{Y_s \mu_s}$, where λ_s is the average number of daily arrivals, and one can recover the performance measures for such system making use of standard queuing theory results (e.g., Hillier & Lieberman, 2014).

The average queue length L_{qs} for each specialty can be computed using the following equations, borrowed from (Hillier & Lieberman, 2014):

$$L_{qs} = \sum_{n=Y_s}^{\infty} (n - Y_s) \pi_n(Y_s), \quad (16)$$

where

$$\pi_n(Y_s) = \begin{cases} \frac{\left(\frac{\lambda_s}{\mu_s}\right)^n}{n!} \pi_0, & 0 \leq n \leq Y_s \\ \frac{\left(\frac{\lambda_s}{\mu_s}\right)^n}{s! s^{n-t}} \pi_0, & n \geq Y_s \end{cases}, \quad (17)$$

and

$$\pi_0(Y_s) = \left(\sum_{n=0}^{Y_s-1} \frac{\left(\frac{\lambda_s}{\mu_s}\right)^n}{n!} + \frac{\left(\frac{\lambda_s}{\mu_s}\right)^{Y_s}}{Y_s!} \cdot \frac{1}{1 - \frac{\lambda_s}{\mu_s}} \right)^{-1}, \quad (18)$$

where $\pi_n(Y_s)$ is the long term probability of having n patients of specialty s in the system, given that Y_s beds are made available for this specialty. In particular $\pi_0(Y_s)$ is the probability that there are no patients of specialty s either waiting for surgery or in the recovery ward.

One can also easily obtain the average number of patients L_s of specialty s in the system and the average waiting times in line for patients of this specialty, making use of the expressions below:

$$L_s = L_{qs} + \frac{\lambda_s}{\mu_s}, \quad (19)$$

$$W_{qs} = \frac{L_{qs}}{\lambda_s}. \quad (20)$$

The average time a patient of specialty s spends in the system is

$$W_s = W_{qs} + \frac{1}{\mu_s}. \quad (21)$$

Such measures are presented in Table 5, where ρ is dimensionless, λ_s and μ_s are given in patients per day Y_s is the number of available beds, L_s and L_{qs} are numbers of patients and W_s and W_{qs} are waiting times in days. Note that the average waiting times, after the optimization problem is solved, are moderately small.

Another measure of interest is the actual waiting time for surgery. One can always use average waiting times to make decisions, but for healthcare problems perhaps it makes more sense to think about the probability of a patient having to wait more than a prescribed time t for surgery. That probability depends on the occupation rate ρ and is given by equation (22).

$$P(W_{qs} > t) = (1 - P(W_q = 0)) e^{-Y_s \mu_s (1 - \rho_s) t}, \quad (22)$$

where

$$P(W_q = 0) = \sum_{n=0}^{Y_s-1} \pi_0(Y_s). \quad (23)$$

These probabilities are shown, in more detail, in Fig. 7 for the specialties of pediatrics and microsurgery, respectively. These results are relative to the open block schedule in Section 5. Observe that the probability that a pediatric patient waits less than 13 days is 0.95. In addition, one can expect 97, 5% and 99% of the patients

Table 5
Performance measures for INTO's surgery queues.

Surgical center	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s
Pediatric	1.96	0.24	9	0.91	6.9166	3.5289	15.0833	7.6955
Feet and Ankle	2.94	0.24	14	0.88	3.7426	1.2730	15.9926	5.4397
Hand	5.98	0.45	18	0.74	0.4566	0.0764	13.7455	2.2986
Hip	4.24	0.09	56	0.84	0.7758	0.1830	47.8870	11.2941
Knee	6.47	0.16	44	0.92	5.4383	0.8405	45.8758	7.0905
Spine	2.63	0.08	40	0.82	0.7573	0.2880	33.6323	12.7880
Shoulder and Elbow	2.41	0.31	17	0.46	0.0025	0.0011	7.7767	3.2269
Microsurgery	0.99	0.17	6	0.97	30.3559	30.6626	36.1795	36.5449
Limb Lengthening	0.75	0.16	7	0.67	0.5156	0.6875	5.2031	6.9375
Oncology	1.29	0.23	9	0.62	0.2359	0.1829	5.8446	4.5307
Skull–Maxillo–Facial	1.34	0.32	13	0.32	0.0002	0.0002	4.1877	3.1252
Adult Trauma	2.02	0.11	28	0.66	0.0471	0.0233	18.4108	9.1142
Elderly Trauma	0.55	0.08	12	0.57	0.0749	0.1362	6.9499	12.6362

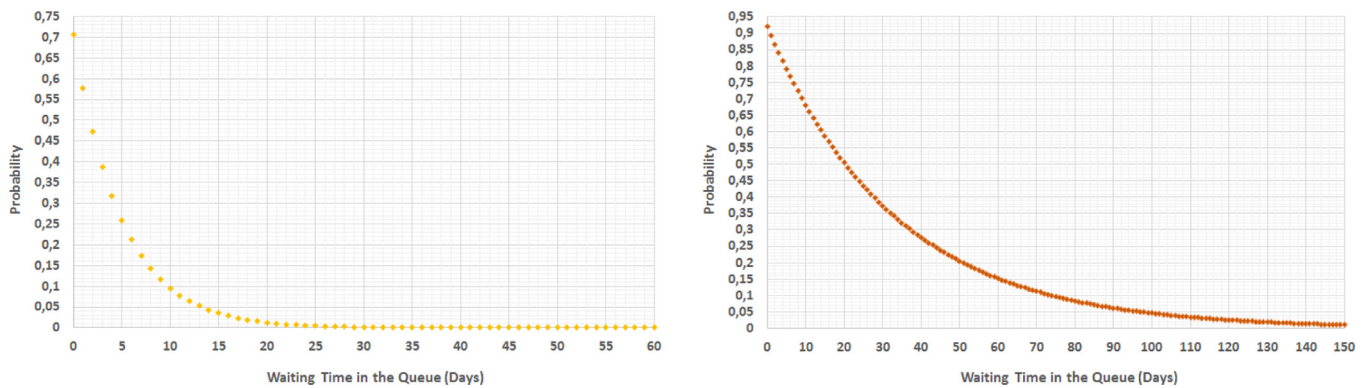


Fig. 7. Waiting time probabilities for pediatric surgeries (left) and microsurgery (right).

to wait less than 16 and 21 days, respectively. Note also that the probability that a patient of the *microsurgery* center waits less than ninety-seven days in line is 95%. Furthermore, about 97, 5% and 99% of the patients will wait less than 120 and 150 days, respectively.

5.4. Other instances and analysis of different scenarios

The analysis of INTO's results, presented in Section 5.2, was important to project the behavior of the queues for each specialty in the future. However, as mentioned previously, a full implementation at INTO is not an easy task.

Therefore, in order to better evaluate the model in this work, we built five hypothetical instances motivated by the six more representative specialties in terms of the required number of recovery ward beds, according to the results in Table 2. They are Hip, Knee, Spine, Adult Trauma, Hand, and Shoulder and Elbow; together, they account for approximately 70% of the beds.

The first instance, called *Scenario 1*, was created by randomly selecting input rates (in number of patients per day), surgery times (in hours), and recovery times (in days) from the range of values of these parameters for the six specialties named above. The other scenarios were created from Scenario 1, each new scenario with a 10% increase in the patient input rate with respect to the previous one, and the resulting parameters are presented in Tables 6 and 7. The aim was to evaluate the effects of the demand increases in the utilization of the surgical center. The resources such as number of beds and number of surgical centers are fixed for all scenarios: there are 230 beds and 14 surgery rooms compatible with every specialty. Moreover, setup and cleanup activities together account for half an hour, regardless of the specialty.

Both closed and open block approaches were tested for all the five hypothetical scenarios. Figs. 8–12 show these results; compar-

Table 6
Daily input rate by specialty.

Surgical center	Scenario 1	Scenario 2	Scenario 3	Scenario 4	Scenario 5
Speciality A	4.73	5.20	5.68	6.15	6.62
Speciality B	1.45	1.60	1.74	1.89	2.03
Speciality C	3.44	3.78	4.13	4.47	4.82
Speciality D	4.63	5.09	5.56	6.02	6.48
Speciality E	1.12	1.23	1.34	1.46	1.57
Speciality F	5.60	6.16	6.72	7.38	7.84

Table 7
Surgery time and recovery time by specialty.

Surgical center	Surgery time (hours)	Recovery time (days)
Speciality A	2.08	4.08
Speciality B	3.21	5.36
Speciality C	3.52	9.98
Speciality D	2.35	6.80
Speciality E	1.85	3.71
Speciality F	2.26	8.92

ing both approaches, one can see that the open block approach makes a better use of the available capacity for all the scenarios.

Fig. 13 summarizes the previous results for all scenarios. A careful observation of Fig. 13 yields that there is a greater efficiency in the utilization of the surgical center when one employs the open block approach. This greater efficiency occurs both in the number of surgeries performed and in the occupation of the surgical center. It is noteworthy that for both approaches, all operating rooms are utilized.

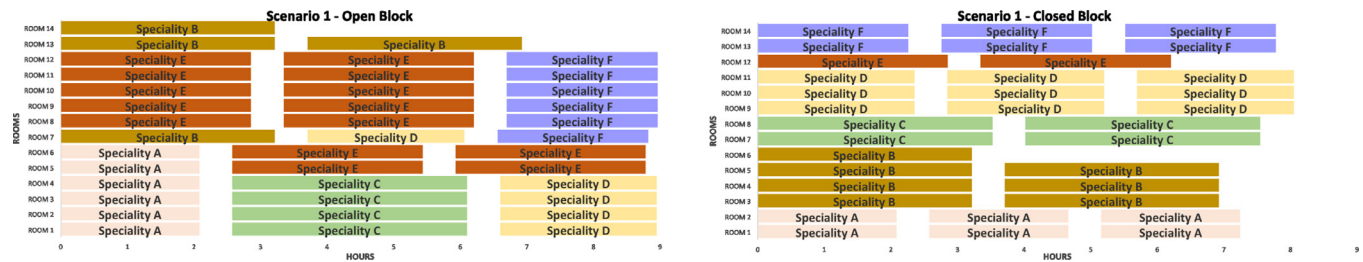
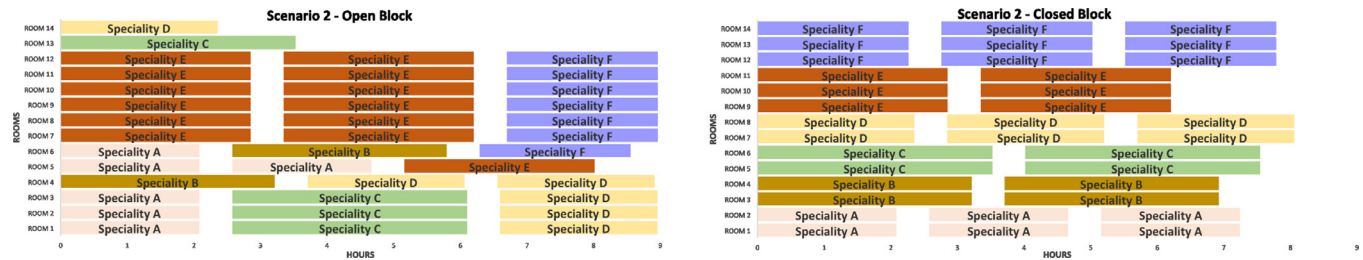
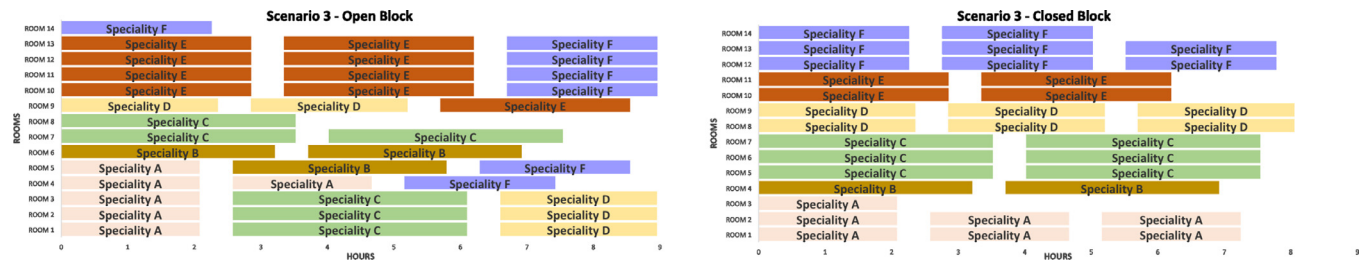
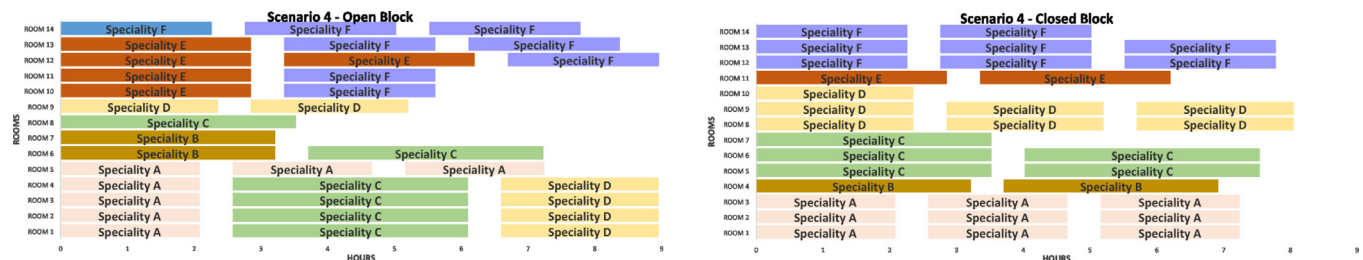
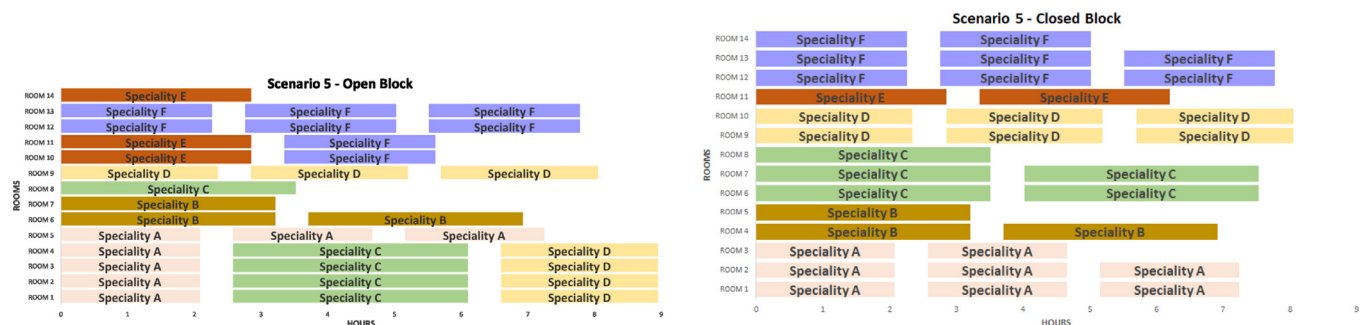
Fig. 8. Results for *Scenario*₁ – open block (left) and closed block (right) approaches.Fig. 9. Results for *Scenario*₂ – open block (left) and closed block (right) approaches.Fig. 10. Results for *Scenario*₃ – open block (left) and closed block (right) approaches.Fig. 11. Results for *Scenario*₄ – open block (left) and closed block (right) approaches.Fig. 12. Results for *Scenario*₅ – open block (left) and closed block (right) approaches.

Fig. 13 also evidences relative decreases in both occupation rates and number of surgeries as the patient input rates increase. That happens because smaller input rates mean that the algorithm is able to assign surgeries in order to simultaneously satisfy constraints (10)–(12) and still be left with spaces in the surgery rooms and the recovery beds, which can be occupied in such a way as to

relatively increase both the surgery room occupation and the number of surgeries. As the demands increase, however, simultaneously satisfying constraints (10)–(12) becomes increasingly harder; hence, the algorithm is left with less room to maneuver, which results in relatively smaller values for both occupation rates and number of surgeries.

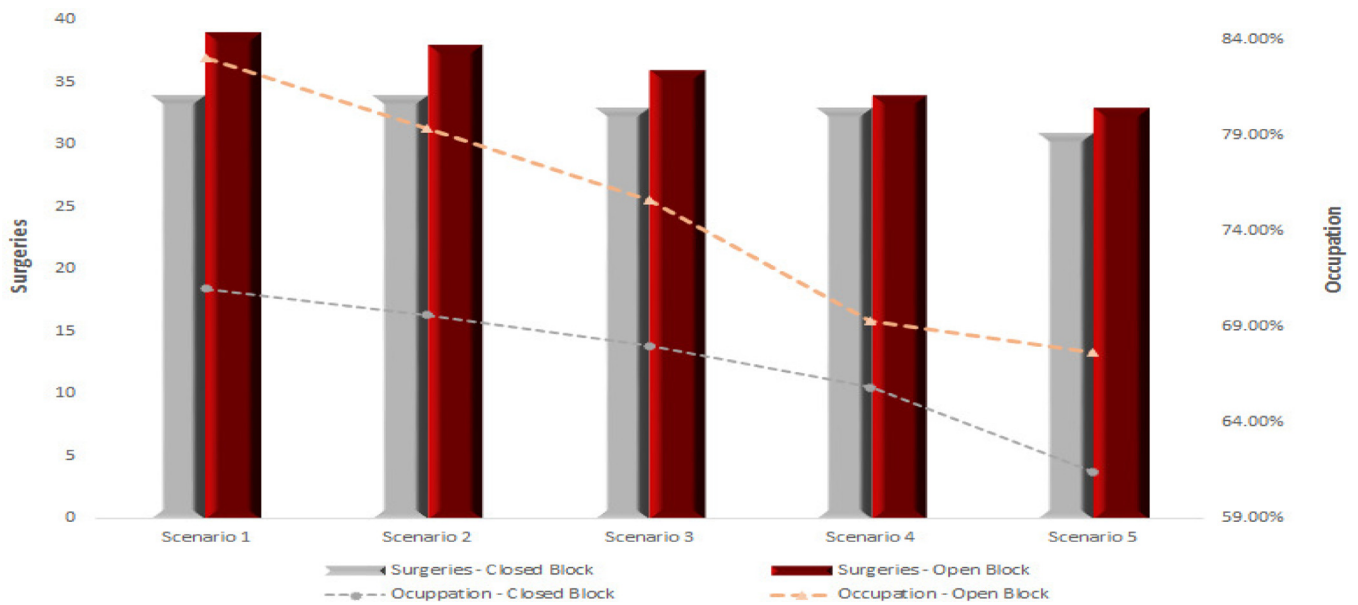


Fig. 13. Scenario analysis: number of surgeries versus occupation surgical center.

Table 8

Performance measures for Scenario 1 surgery queues – open block (left) and closed block (right) approaches.

Specialty	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s	Specialty	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s
Specialty A	4.73	0.25	25	0.7568	0.4065	0.086	19.3265	4.086	Specialty A	4.73	0.25	25	0.7568	0.4065	0.086	19.3265	4.086
Specialty B	1.45	0.19	24	0.318	0	0	7.6316	5.2632	Specialty B	1.45	0.19	40	0.1908	0	0	7.6316	5.2632
Specialty C	3.44	0.1	41	0.839	1.043	0.3032	35.443	10.3032	Specialty C	3.44	0.1	41	0.839	1.043	0.3032	35.443	10.3032
Specialty D	4.63	0.15	34	0.9078	4.7192	1.0193	35.5859	7.6859	Specialty D	4.63	0.15	62	0.4978	0	0	30.8667	6.6667
Specialty E	1.12	0.27	52	0.0798	0	0	4.1481	3.7037	Specialty E	1.12	0.27	8	0.5185	0.076	0.0678	4.2241	3.7715
Specialty F	5.6	0.11	54	0.9428	9.4029	1.6791	60.312	10.77	Specialty F	5.6	0.11	54	0.9428	9.4029	1.6791	60.312	10.77

Table 9

Performance measures for Scenario 2 surgery queues – open block (left) and closed block (right) approaches.

Specialty	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s	Specialty	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s
Specialty A	5.2	0.25	25	0.832	1.4144	0.272	22.2144	4.272	Specialty A	5.2	0.25	27	0.7704	0.4648	0.0894	21.2648	4.0894
Specialty B	1.6	0.19	11	0.7656	1.0292	0.6432	9.4502	5.9064	Specialty B	1.6	0.19	18	0.4678	0.0026	0.0016	8.4236	5.2648
Specialty C	3.78	0.1	40	0.945	10.9062	2.8852	48.7062	12.8852	Specialty C	3.78	0.1	40	0.945	10.9062	2.8852	48.7062	12.8852
Specialty D	5.09	0.15	42	0.8079	0.5353	0.1052	34.4687	6.7718	Specialty D	5.09	0.15	41	0.8276	0.8279	0.1627	34.7613	6.8293
Specialty E	1.23	0.27	49	0.093	0	0	4.5556	3.7037	Specialty E	1.23	0.27	23	0.1981	0	0	4.5556	3.7037
Specialty F	6.16	0.11	63	0.8889	2.1467	0.3485	58.1467	9.4394	Specialty F	6.16	0.11	81	0.6914	0.0024	0.0004	56.0024	9.0913

Table 10

Performance measures for Scenario 3 surgery queues – open block (left) and closed block (right) approaches.

Specialty	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s	Specialty	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s
Specialty A	5.68	0.25	25	0.9088	5.4119	0.9528	28.1319	4.9528	Specialty A	5.68	0.25	30	0.7573	0.3171	0.0558	23.0371	4.0558
Specialty B	1.74	0.19	17	0.5387	0.0168	0.0096	9.1747	5.2728	Specialty B	1.74	0.19	11	0.8325	2.3155	1.3307	11.4734	6.5939
Specialty C	4.13	0.1	50	0.826	0.6303	0.1526	41.9303	10.1526	Specialty C	4.13	0.1	60	0.6883	0.0089	0.0022	41.3089	10.0022
Specialty D	5.56	0.15	41	0.9041	3.9948	0.7185	41.0615	7.3852	Specialty D	5.56	0.15	41	0.9041	3.9948	0.7185	41.0615	7.3852
Specialty E	1.34	0.27	34	0.146	0	0	4.963	3.7037	Specialty E	1.34	0.27	15	0.3309	0.0001	0.0001	4.9631	3.7038
Specialty F	6.72	0.11	63	0.9697	23.5227	3.5004	84.6136	12.5913	Specialty F	6.72	0.11	72	0.8485	0.6714	0.0999	61.7623	9.1908

Tables 8–12 present the performance measures of the surgery queues for all five scenarios, under the same conditions applied in Section 5.3, namely Poisson patient arrival processes and exponential recovery times. Note that the most congested queue is that of Specialty F in Scenario 5, with $\rho \approx 0.99$ in both closed and open block approaches. Fig. 14 features the waiting time distribution for this queue; it shows that the 95% of the patients wait up to 36 days for surgery, whereas 99% of them wait less than 55 days.

6. Concluding remarks

Motivated by a case study at the Brazilian Institute of Traumatology and Orthopedics (INTO), this paper introduces an integrated model for long-term planning and scheduling of elective surgeries. The models addresses the whole process, from the patient arrival to their release from the recovery ward. Considering that only an integrated approach can promote a consistent balance between service and demand, our approach produces both a periodic

Table 11

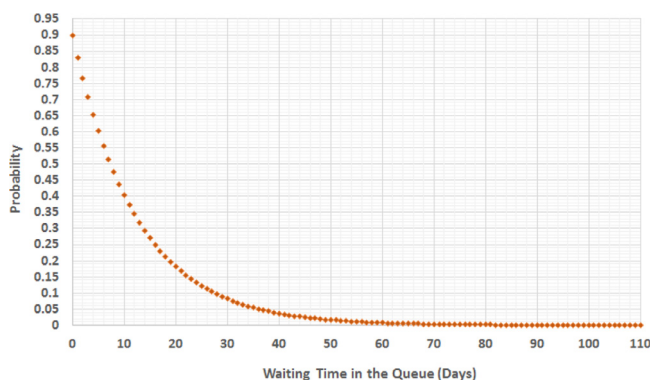
Performance measures for Scenario 4 surgery queues – open block (left) and closed block (right) approaches.

Specialty	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s	Specialty	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s
Specialty A	6.15	0.25	29	0.8483	1.67	0.2715	26.27	4.2715	Specialty A	6.15	0.25	37	0.6649	0.0259	0.0042	24.6259	4.0042
Specialty B	1.89	0.19	11	0.9043	6.3051	3.336	16.2525	8.5992	Specialty B	1.89	0.19	11	0.9043	6.3051	3.336	16.2525	8.5992
Specialty C	4.47	0.1	50	0.894	2.8549	0.6387	47.5549	10.6387	Specialty C	4.47	0.1	54	0.8278	0.593	0.1327	45.293	10.1327
Specialty D	6.02	0.15	48	0.8361	0.8286	0.1376	40.9619	6.8043	Specialty D	6.02	0.15	48	0.8361	0.8286	0.1376	40.9619	6.8043
Specialty E	1.46	0.27	20	0.2704	0	0	5.4074	3.7037	Specialty E	1.46	0.27	8	0.6759	0.4882	0.3344	5.8956	4.0381
Specialty F	7.28	0.11	72	0.9192	4.295	0.59	70.4769	9.6809	Specialty F	7.28	0.11	72	0.9192	4.295	0.59	70.4769	9.6809

Table 12

Performance measures for Scenario 5 surgery queues – open block (left) and closed block (right) approaches.

Specialty	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s	Specialty	λ_s	μ_s	Y_s	ρ_s	L_{qs}	W_{qs}	L_s	W_s
Specialty A	6.62	0.25	29	0.9131	5.6022	0.8462	32.0822	4.8462	Specialty A	6.62	0.25	28	0.9457	12.0541	1.8209	38.5341	5.8209
Specialty B	2.03	0.19	22	0.4856	0.0016	0.0008	10.6858	5.2639	Specialty B	2.03	0.19	16	0.6678	0.1875	0.0924	10.8717	5.3555
Specialty C	4.82	0.1	50	0.964	19.3284	4.0101	67.5284	14.0101	Specialty C	4.82	0.1	50	0.964	19.3285	4.0101	67.5284	14.0101
Specialty D	6.48	0.15	49	0.8816	2.1942	0.3386	45.3942	7.0053	Specialty D	6.48	0.15	48	0.9	3.3538	0.5176	46.5538	7.1842
Specialty E	1.57	0.27	8	0.7269	0.8388	0.5343	6.6536	4.238	Specialty E	1.57	0.27	16	0.3634	0.0002	0.0001	5.815	3.7038
Specialty F	7.84	0.11	72	0.9899	88.1325	11.2414	159.4052	20.3323	Specialty F	7.84	0.11	72	0.9899	88.1325	11.2414	159.4052	20.3323

**Fig. 14.** Waiting time probabilities for Specialty F, under Scenario 5.

surgery scheduling plan, called master schedule, and an optimal bed allocation policy for the recovery ward. Both the master schedule and the bed allocation policy are produced in such a way that the long term balance between offer and demand is guaranteed for all surgical specialties.

The paper also proposes a queuing analysis of the stochastic process resulting from the application of the prescribed master schedule and bed allocation plans. Such an analysis provides the decision maker with a better understanding of the queues for all surgical specialties, allowing decisions to be taken in terms of significant queuing measures, such as the average waiting time or the maximum waiting time for a prescribed percentage of patients.

Acknowledgments

The authors thank the editor and the reviewers for their invaluable contributions, which resulted in a much improved text.

The authors would like to thank the Carlos Chagas Filho Foundation for Research Support of the State of Rio de Janeiro, FAPERJ, for supporting the research by means of Grant no. E-26/202.789/2015. This work was partially supported by the National Council for Scientific and Technological Development – CNPq, under Grants 303543/2015-9 and 308978/2013-7.

References

- Agnetis, A., Coppi, A., Corsini, M., Dellino, G., Meloni, C., & Pranzo, M. (2014). A decomposition approach for the combined master surgical schedule and surgical case assignment problems. *Health Care Management Science*, 17(1), 49–59. doi:10.1007/s10729-013-9244-0. ISSN 1386-9620

- Aringhieri, R., Landa, P., Soriano, P., Tanfani, E., & Testi, A. (2015a). A two level meta-heuristic for the operating room scheduling and assignment problem. *Computers and Operations Research*, 54, 21–34. doi:10.1016/j.cor.2014.08.014.
- Aringhieri, R., Landa, P., & Tanfani, E. (2015b). Assigning surgery cases to operating rooms: A VNS approach for leveling ward beds occupancies. *Electronic Notes in Discrete Mathematics*, 47, 173–180. doi:10.1016/j.endm.2014.11.023.
- Augusto, V., Xie, X., & Perdomo, V. (2010). Operating theatre scheduling with patient recovery in both operating room and recovery beds. *Computers and Industrial Engineering*, 58(2), 231–238.
- Belien, J., & Demeulemeester, E. (2007). Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research*, 176(2), 1185–1204.
- Belien, J., & Demeulemeester, E. (2008). A branch-and-price approach for integrating nurse and surgery scheduling. *European Journal of Operational Research*, 189(3), 652–668.
- Bhattacharjee, P., & Ray, P. K. (2014). Patient flow modeling and performance analysis of healthcare delivery processes in hospitals: An review and reflections. *Computers and Industrial Engineering*, 78, 299–312.
- Blake, J. T. (2010). Capacity planning in operating rooms. In Y. Yih (Ed.), *Handbook of healthcare delivery systems* (pp. 34–1–34–12). Boca Raton: CRC Press. ISBN 9781439803332 - CAT# K10187
- Bowers, J., & Mould, G. (2004). Managing uncertainty in orthopaedic trauma theatres. *European Journal of Operational Research*, 154(3), 599–608.
- Cardoen, B., Demeulemeester, E., & Belien, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Research*, 201(3), 921–932.
- Demeulemeester, R., Belien, J., Cardoen, B., & Samudra, M. (2013). Operating room planning and scheduling. In B. Denton (Ed.), *Handbook of healthcare operations management*. In *International series in operations research and management science*: 184 (pp. 121–152). New York: Springer. doi:10.1007/978-1-4614-5885-2_5. ISBN 978-1-4614-5884-5
- Hans, E. W., & Vanberkel, P. T. (2012). Operating theatre planning and scheduling. In R. Hall (Ed.), *Handbook of healthcare system scheduling*. In *International series in operations research and management science*: 168 (pp. 105–130). New York: Springer.
- Hillier, F. S., & Lieberman, G. J. (2014). *Introduction to operations research* (10th). New York: McGraw-Hill.
- Jonnalagadda, R., Hariharan, E. R., Walrond, S., & Prasad, C. (2005). Evaluation of the reasons for cancellations and delays of surgical procedures in a developing country. *International Journal of Clinical Practice*, 59(6), 716–720.
- Kleinrock, L. (1975). *Queueing systems, volume 1: Theory*. New York: John Wiley & Sons.
- Liu, Y., Chu, C., & Wang, K. (2011). A new heuristic algorithm for the operating room scheduling. *Computers and Industrial Engineering*, 61, 865–871.
- Marques, I., Captivo, M. E., & Pato, M. V. (2012). An integer programming approach to elective surgery scheduling. *OR Spectrum*, 34, 407–427.
- Milne, E., & Whitty, P. (1995). Calculation of the need for paediatric intensive care beds. *Archives of Disease in Childhood*, 73(6), 505–507.
- Min, D., & Yih, Y. (2010a). An elective surgery scheduling problem considering patient priority. *Computers and Operations Research*, 37(6), 1091–1099.
- Min, D., & Yih, Y. (2010b). Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research*, 206, 642–652.
- Molina-Pariente, J. M., Fernandez-Viagas, V., & Framinan, J. M. (2015). Integrated operating room planning and scheduling problem with assistant surgeon dependent surgery durations. *Computers and Industrial Engineering*, 82, 8–20. doi:10.1016/j.cie.2015.01.006.
- Pearson, G., Reynolds, F., & Stickley, G. (2012). Calculating the need for intensive care beds. *Archives of Disease in Childhood*, 97(11), 943–946. doi:10.1136/archdischild-2011-301310.

- Pham, D., & Klinkert, A. (2008). Surgical case scheduling as a generalized job shop scheduling problem. *European Journal of Operational Research*, 185(3), 1011–1025.
- Saadouli, H., Jerbi, B., Dammak, A., Masmoudi, L., & Bouaziz, A. (2015). A stochastic optimization and simulation approach for scheduling operating rooms and recovery beds in an orthopedic surgery department. *Computers and Industrial Engineering*, 80, 72–79.
- Santibanez, P., Begen, M., & 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.
- Tanfani, E., & Testi, A. (2010). A pre-assignment heuristic algorithm for the master surgical schedule problem(MSSP). *Annals of Operations Research*, 178, 105–119.