Annals of Operations Research (2020) 286:501–527
https://doi.org/10.1007/s10479-018-3008-7

S.I.: CLAIO 2016

Operating room scheduling under waiting time constraints: the Chilean GES plan

Javiera Barrera[1] · Rodrigo A. Carrasco[1] · Susana Mondschein[1] · Gianpiero Canessa[1] · David Rojas-Zalazar[2]

Published online: 19 August 2018

Abstract
In 2000, Chile introduced profound health reforms to achieve a more equitable and fairer system (GES plan). The reforms established a maximum waiting time between diagnosis and treatment for a set of diseases, described as an opportunity guarantee within the reform. If the maximum waiting time is exceeded, the patient is referred to another (private) facility and receives a voucher to cover the additional expenses. This voucher is paid by the health provider that had to do the procedure, which generally is a public hospital. In general, this reform has improved the service for patients with GES pathologies at the expense of patients with non-GES pathologies. These new conditions create a complicated planning scenario for hospitals, in which the hospital's OR Manager must balance the fulfillment of these opportunity guarantees and the timely service of patients not covered by the guarantee. With the collaboration of the *Instituto de Neurocirugía*, in Santiago, Chile, we developed a mathematical model based on stochastic dynamic programming to schedule surgeries in order to minimize the cost of referrals to the private sector. Given the large size of the state space, we developed an heuristic to compute *good* solutions in reasonable time and analyzed its performance. Our experimental results, with both simulated and real data, show that our algorithm performs close to optimum and improves upon the current practice. When we compared the results of our heuristic against those obtained by the hospital's OR manager in a simulation setting with real data, we reduced the overtime from occurring 21% of the time to zero, and the non-GES average waiting list's length from 71 to 58 patients, without worsening the average throughput.

Keywords Scheduling · Operating theater · Operating room scheduling

1 Introduction

In 2015, the World Health Assembly unanimously passed a resolution highlighting the critical role of essential surgical and anesthesia care in achieving universal health coverage, stressing the importance of access to timely, safe, and affordable surgical and anesthesia services

✉ Susana Mondschein
susana.mondschein@uai.cl

Extended author information available on the last page of the article

(World Health Organization 2017). But health care providers around the world have faced significant increases in waiting times for surgery and other medical services. If treatment is not provided in a medically reasonable time, then these excessive delays can deteriorate the recovery process. To prevent this problem, health authorities have committed to determining the maximum waiting times, also referred as *due dates*, for a critical subset of diagnoses.

Due dates can be determined by the surgery team, as in Vansteenkiste et al. (2012), or by the health authority, as in Patrick and Puterman (2008). Independently of who determines these performance indicators, in the last 15 years, waiting times and other related metrics have become significant in operating room (OR) planning [see, for example, Figure 3 in Samudra et al. (2016)].

In 2000, Chile introduced profound health reforms that aimed to achieve a more equitable and fairer system (Lenz 2007), with a law enacted in August 2004. This reform, initially known as the AUGE Plan and later renamed GES, consists of guarantees regarding access, quality, opportunity, and financial protection for all Chilean citizens diagnosed with a pathology within an established set of diagnoses (currently in 80 pathologies). If a patient is diagnosed with one of these pathologies, she is officially notified of her rights as a GES patient and is entered into a centralized managed database that tracks all subsequent medical attentions.

In particular, the opportunity guarantee specifies a maximum number of days for the patient's waiting time before receiving appropriate treatment. If the opportunity guarantee cannot be achieved in due time for any reason, the insurer has two days to identify an alternative provider (public or private) that is able to fulfill it. This new provider then has 10 days to deliver the appropriate treatment guaranteed by law. The financial protection guarantee remains for the patient; therefore, the potential cost difference is incurred by the original health provider. Thus, the patient receives a voucher identifying the health service required and the newly assigned health provider.

The implementation of the GES plan has proven to be a challenge for OR managers, who are pressured to provide certain health treatments within fixed time ranges, possibly postponing other health services that are not included in the 80 pathologies covered by the GES plan. For example, according to a 2014 report of the Ministry of Health, 56% of non-GES pathologies have been waiting for more than a year for surgery; a total of 186,377 non-GES patients were not treated between 2009 and 2014 (Catalina De Améstica 2014).

Therefore, the hospital manager not only has to address the extra financial burden represented by the cost of GES vouchers when the time guarantee is not satisfied, but also has to consider excessive delays in patient treatment, which result in patient dissatisfaction and, more importantly, negative health consequences. This is especially true for pathologies not covered by the GES plan, whose patients have seen a worsening in service level with the implementation of the new system (Zúñiga-Fajuri 2007).

OR managers regularly face the problem of allocating shared OR time slots among services including traumatology, gynecology, oncology, and neurology while attempting to satisfy different stakeholders (mostly OR managers). Each manager handles her assigned OR time and allocates it among different types of surgeries, again attempting to satisfy different stakeholders such as patients, surgeons, and anesthesiologists. For example, surgeons and anesthesiologists often request a stable workload over time and specific shifts during the day and week. On the other hand, patients appreciate on-time and high-quality service, see Section 4 in Cardoen et al. (2010) for a discussion of the performance metrics and their relationships with the different stakeholders.

In this paper, we study the problem faced by a OR manager, who must assign available OR capacity among patients, while considering the extra cost incurred when the GES opportunity guarantee is not satisfied. This study was conducted at the Instituto de Neurocirugía Dr. A. Asenjo in Santiago, Chile. This hospital is a specialty facility, where patients with neurosurgical needs, from around the country, are treated. Therefore, the strategic and tactical decision levels, where the available operating room time is divided over the different surgical groups with the corresponding master surgery schedule, typically a cyclic schedule approach, are not required. Our study focuses at the operational level where the detailed weekly and/or daily scheduling takes place, including assigning the specific patient/surgeons to operating rooms, and determining the order and the start and end times of the surgeries.

The Instituto de Neurocirugía has four operating rooms, where more than 120 different types of surgeries are performed, with approximately 460 monthly procedures. A detailed list of the 14 surgeries that account for more than 80% of the OR utilization time can be found in Appendix A.1. The current practice for the scheduling process is as follows: on Thursdays, a medical team composed by two physicians, a nurse, and three resident doctors, revises the waiting list and schedule patients for the following week. Based on the team's experience, GES surgeries with opportunity guarantees that expire during the next 2 weeks are scheduled first. Then, non-GES patients are scheduled for the week, with priority given to those who have been waiting the longest and/or have a higher risk of complications. This process consumes approximately 2 h of time for the six health care professionals. The scheduling process is manual and relies heavily on the personal experience of the scheduling team, to the point that the main two physicians are still consulted on the procedure when they are away in a conference or vacations. The latter provides an opportunity not only to optimize the scheduling procedure in terms of performance measures but also to save valuable professional resources.

In this work, we formulate this problem as a stochastic dynamic programming (SDP) model, in order to minimize the expected cost of GES vouchers, taking into account a minimum service level for non-GES patients. An important feature of our model is the consideration of physician-dependent surgery times, which is reported as important in the literature (Stepaniak et al. 2010) and is further confirmed in our collected data.

Given the dimensions of the state space of our problem, which makes solving the SDP to optimality infeasible for meaningful instances, we develop an heuristic, named GES-PROG, that provides good solutions in a reasonable computational time. We then compare the solutions given by GES-PROG, with the value of a lower bound for the problem. Our computational experiments show that the GES-PROG performs very close to optimal. We study several settings, through simulated instances, that demonstrate the good performance of the algorithm.

Finally, to validate the usefulness of our algorithm, we applied this methodology to a real case study conducted at the Instituto de Neurocirugía. For this study, we compared the performance of our algorithm with those obtained by the hospital OR manager within a simulation setting. To develop this simulation platform, we collected data from 2015 to 2016 and computed the relevant parameters, such as surgery times, variances, etc. The solutions obtained by our algorithm led to significant improvements in the waiting list and the overtime usage in the OR. In particular, while maintaining a throughput of approximately 30 surgeries per week and a utilization of 90%, the overtime improved from having 21% of days with overtime to zero. It is important to note that the average overtime is 1 h, when there was overtime. Additionally, the average waiting list's length was reduced by our algorithm in 20% from 71 to 58 non-GES patients.

Our contributions The main contributions of our work are as follows:

- We consider due dates as the ones imposed by the health authorities for the GES subset of pathologies, while warranting capacity for non-GES ones. Furthermore, unlike the previous literature, these due dates are not mandatory but impose a step-wise type of cost function when surgery due dates are not met.
- In this study we are interested in the OR planning problem that considers only elective surgeries, but incorporating the surgery time dependence not only on the pathology but also on the physician.
- We develop a heuristic to compute solutions quickly, which we further compare to an analytical lower bound showing that the computed solutions are very good.
- We implemented and tested our algorithm in a real case setting, comparing it to the OR Manager's solution. The solutions obtained by our algorithm led to significant improvements in the waiting list and the overtime utilization of the OR in the context of the hospital that motived our study.

The remainder of this paper is organized as follows: Sect. 2 reviews the existing literature. Sections 3 and 4 provide the model formulation and the corresponding heuristics. Section 5 details the input data used in the simulation-based performance analysis in Sect. 6 and in the performance validation in Sect. 7. Finally, conclusions and future research lines are discussed in Sect. 8.

2 Literature review

There is a wide body of literature on the topic of managing health care capacity. This problem has been studied using quantitative tools since the early 1960s, providing decision makers with a wide range of solutions. Demand for health care services has steadily increased over time due to: (i) the emigration of populations from rural to urban areas in low- and middle-income countries, and (ii) the increased aging population in high-income countries (Brandeau et al. 2004). Therefore, health providers have been pressured to efficiently use resources without diminishing patient satisfaction. The operating room is one of the most expensive units in a hospital, and according to Denton et al. (2007), it accounts for 40% of the total expenses. There is a vast amount of literature regarding the scheduling problem of operating rooms. This array of approaches arose because, although all OR managers face under-/over-utilization of the ORs, they must also consider different performance metrics, stakeholders and how they interact, and other specific hospital features. Surveys Cardoen et al. (2010) and Guerriero and Guido (2011) reviewed more than 150 articles on this topic. The former survey provides an updated overview of OR planning and scheduling that captures recent developments while the latter focuses on how operations research can be applied to OR scheduling to balance the interests of different stakeholders.

The OR planning problem can be studied at the strategic, tactical, and operational levels. Although decisions in these three stages are highly interrelated, they are studied separately because of their complexity and different time horizons [see, for example, Beliën and Demeulemeester (2007) and Santibáñez et al. (2007)]. At the strategic level, hospitals assign the OR capacity to the different specialties, such as, for example, traumatology, gastroenterology and cardiology, in order to balance out the different needs of the medical teams. Then, at the tactical level, and once the strategic capacity allocation is solved, the hospitals determine which ORs, days and block of hours are assigned to each specialty. Typically, the hospitals solve this assignment by building a *Block Schedule*, also called a *Master Surgical Schedule or OR Block Allocation Table*, that usually changes every six to twelve months. Finally, at the operational level, each specialty determines a daily assignment of patients–physicians to specific operating rooms and starting times. This problem is usually divided into two steps: (i) the process of fixing a surgery date for a patient, known as *advance scheduling*, and (ii) the process of determining the operating room and the starting time of the procedure on the specific date of surgery, known as *allocation scheduling* (Cardoen et al. 2010; Magerlein and Martin 1978). Notice that, at this level, these decisions can significantly differ depending on the administrative practice of the hospital and whether or not the OR capacity is shared between elective and non-elective surgeries. For example, planning horizons may extend from one week to a month and OR planning may include one or several sub-specialties, or consider downstream units such as an intensive care unit (ICU) or post-anesthesia care unit (PACU).

As mentioned in the Introduction, the Instituto de Neurocirugía Dr. A. Asenjo, where we have applied this study, only does neurosurgical procedures. It has one OR dedicated to emergency surgeries and four ORs dedicated to elective surgeries: three for adult surgeries and one shared between adult and pediatric procedures (adults on Mondays and children from Tuesdays to Fridays). Thus, the Block Schedule assignment is given, and therefore, we focus at the operational level considering two important features: (i) the surgery time depends not only on the pathology but also on the physician performing the procedure, and (ii) GES pathologies have a due date, which when not satisfied, the patient is referred to a private provider incurring in a higher cost. In this setting, the OR manager chooses the patients, from the waiting list, to be scheduled for surgery for the upcoming week, based on the availability of physicians, anesthesiologists and OR capacity. This schedule considers that failure to satisfy the due time for GES diagnosis will result in additional referral costs for the hospital. Currently, the OR planning is conducted using spreadsheets and relies on the experience of the OR team.

In our formulation, the OR capacity is shared between GES and non-GES patients. Thus, certain amount of capacity should be guaranteed to non-GES procedures. Otherwise, these could arbitrarily be postponed each time a GES patient needs surgery. Additionally, in our formulation, physicians and anesthesiologists are scheduled for several weeks. Given the scope of our problem, we focus on the OR planning literature that addresses three aspects: the patients' waiting time costs, the *advance scheduling* decisions of multiple ORs, and the variability of surgery times.

Among the articles reviewed that include due dates for surgeries, several consider the patients' waiting time as the key performance measure when making operational decisions [see, for example, Samudra et al. (2016), and references therein]. Others consider the patients' priority combined with the waiting time. In Patrick and Puterman (2008) and Molina and Framinan (2009) they consider mandatory due dates, or equivalently, an infinite cost each time the due date is exceeded. More precisely in Patrick and Puterman (2008) patients are categorized in different priority classes, where each class has a *service time target* (due date). In this article an optimal scheduling policy is proposed to schedule patients on a single resource using queueing theory. The authors proposed a set of criteria to enable schedulers to book a single OR in order to meet the waiting time targets for all classes. In Molina and Framinan (2009), they focus, however, on the difference between assigning a patient to an OR and then to a surgeon versus the other way around, under the due dates constraints. Other studies consider desirable, but not mandatory, due dates. In these cases, a cost, proportional to the deviation time from the due date (extra time), is incurred. In Testi and Tanfani (2009), this cost was used in a binary-linear programming model that, simultaneously, solves a dynamic block problem and the *advance scheduling*, while minimizing the overall patients welfare loss. In Aringhieri et al. (2015) the authors address the *Block* and *Advance Scheduling* simultaneously in a binary-linear programming formulation, exploiting the stated hierarchy of both decision levels. They use an objective function that grows linearly with the waiting time, in order to minimize the total cost of patients' waiting time at the end of the planning horizon. A similar approach is used in Tànfani and Testi (2010) to obtain a *Block schedule* that reduces the weighted waiting times simultaneously with improving the hospital efficiency.

In this paper, we focus at the operational decision level, by solving the *advance* and *allocation scheduling*, considering that the *Block Schedule* is given. Several papers have considered this approach, but differ in capturing the particular setting of our study. For example, in Jebali et al. (2006) the objective function minimizes the cost incurred by keeping the patients in hospital waiting to be treated in the operating room, as well as undertime and overtime costs, indirectly considering patients' waiting time. Another example is Addis et al. (2016) where the *advance scheduling* is generated considering the possibility of a cancellation due to two sources of uncertainty: surgery times, and non-elective patient arrivals. The problem is solved using a rolling horizon in which the mid-term optimization window is moved one week at each iteration. The objective is to minimize an overall penalty that grows linearly with the number of days that surgeries are postponed after the due dates. A similar approach is developed in Astaraky and Patrick (2015), Bruni et al. (2015), Min and Yih (2010), where a stochastic dynamic programming model is proposed for the *advance scheduling* problem. In Astaraky and Patrick (2015), the objective is to minimize a combination of patients' waiting times, overtime in the ORs, and congestion in the wards. In Bruni et al. (2015), the goal is to schedule elective surgeries in a planning horizon of a week, considering simultaneously random service times, random emergency occurrences, and patient priorities by allowing some reasonable overtime. The problem is formulated as a Markov Decision Process, using a version of the Least Squares Approximate to find a policy. Finally in Min and Yih (2010), the objective function captures the notion that higher waiting times may result on higher health care costs, due to additional treatments.

In what follows, we discuss the variability of surgery times. It has been reported since the 70's that the effectiveness of the scheduling depends on the accuracy of the procedure-time estimates. If the estimates are consistently lower than the times actually incurred, the schedule will be overloaded, whereas if the estimates are consistently higher, idle time will result (Magerlein and Martin 1978). There is evidence that surgery times present significant variability for a single procedure. This could be explained by patient attributes, physician practices, type of anesthesia, or random events. This heterogeneity has been reported and analyzed in the recent literature at various hospitals; see, for example, Dexter et al. (2008) and Stepaniak et al. (2010), where the authors suggest that using this information to improve the accuracy of surgery times would improve OR management. In particular, at the Instituto de Neurocirugía Dr A. Asenjo, according to the OR manager, surgery times vary significantly among physicians, and these differences are considered at tactical and operational OR planning. In fact, our analysis of the database showed cases where this difference across physicians is greater than 60% [see Azar et al. (2017) for details]. Most of the studies addressing this uncertainty consider a single probability distribution for surgery time that takes into account all sources of uncertainty. This leads to a probability distribution that only depends on the type of procedure; see for example, Astaraky and Patrick (2015), Min and Yih (2010). A robust formulation for surgery planning is presented in Hans et al. (2008), where the objective is to maximize the capacity utilization and minimize the risk of overtime, assuming that the operating times for each type of surgery are normally distributed. In Bruni et al. (2015), the authors developed a stochastic programming model in which overtime acts as a cushion to absorb variability in operating times and emergency arrivals. In Landa et al. (2016), the authors proposed a chance-constrained formulation to address variability in surgery time and implemented a local search optimization module using Monte Carlo simulation to handle several random surgery times. In Duma and Aringhieri (2015), when the weekly schedule is implemented, if there is a significant delay during a particular day, the OR manager chooses between using overtime or canceling and re-scheduling some of the remaining surgeries. Thus, the OR manager assumes an active role to manage the variability of surgery times. A handful of papers have explicitly considered variability among surgeons to improve the assignment of physicians to patients to increase throughput. This approach was used in Gomes et al. (2012), where data mining techniques were implemented to predict the surgery times: specific physician's operating time estimations were used to maximize the average OR utilization in the scheduling formulation.In our paper, we use a similar approach. We propose a formulation for OR planning where, for a given pair of physician-surgery procedure, a deterministic surgery time is considered. This is estimated using historical information to reduce variability.

In our formulation, we proceed by dividing the operational scheduling into two stages. First, at the *advance scheduling*, the surgery date, surgeon and anesthesiologist are assigned to each patient. We consider a planning horizon of several weeks to take into account the complete waiting list with the corresponding due dates of all patients and future arrivals. Second, at the *allocation scheduling*, we take the assignment for the upcoming week and determine a starting time and date for each patient's surgery. According to the Chilean context, GES health plan imposes due dates for a subset of pathologies. If the due date is exceeded, the hospital incurs in a fixed cost for transferring the patient to a private facility. This cost can be twice to three times the cost of treating the patient in the hospital, and therefore, it impacts, in terms of costs, might be highly significant. It is fundamental, in this case, to capture this behavior, and model the cost as a step function, which we have not found previously in the literature. These types of functions are generally difficult to deal with in scheduling contexts (Pinedo 2012), but in our setting it is necessary to model the requirement that the GES program imposes to the hospital. Finally, in the setting we are studying, emergencies are served in a dedicated OR, and therefore, the main source of variability is from surgery times. We chose to deal with this by reducing the uncertainty through explicitly using the detailed information of surgery times by physician–pathology pairs.

Our work differs from the aforementioned literature in providing an *advance scheduling* that assigns physicians to patients according to their specific surgery times. Moreover, in order to capture the reality of Chilean hospitals, our formulation incorporates the cost of not meeting due dates according to a step function, that introduces a new degree of difficulty compared to those reported in the literature. We remark that these two features are considered in the OR scheduling at the hospital we work with. In our paper, we develop a heuristic to solve the general stochastic dynamic formulation for the weekly scheduling problem, in the presence of multiple ORs. We study the benefits of scheduling physicians–patients according to this setting.

Although it is clear that there are medical reasons to assign a physician to a particular patient, there are patients for whom the choice of physician is flexible. This would lead to a higher OR throughput [see discussion in Molina and Framinan (2009)]. Our formulation and resolution algorithm are flexible to evaluate the increase of throughput under these two scenarios. Finally, we test our algorithm using real data to illustrate the simulation-based performance analysis when the algorithm is used in a real case scenario.

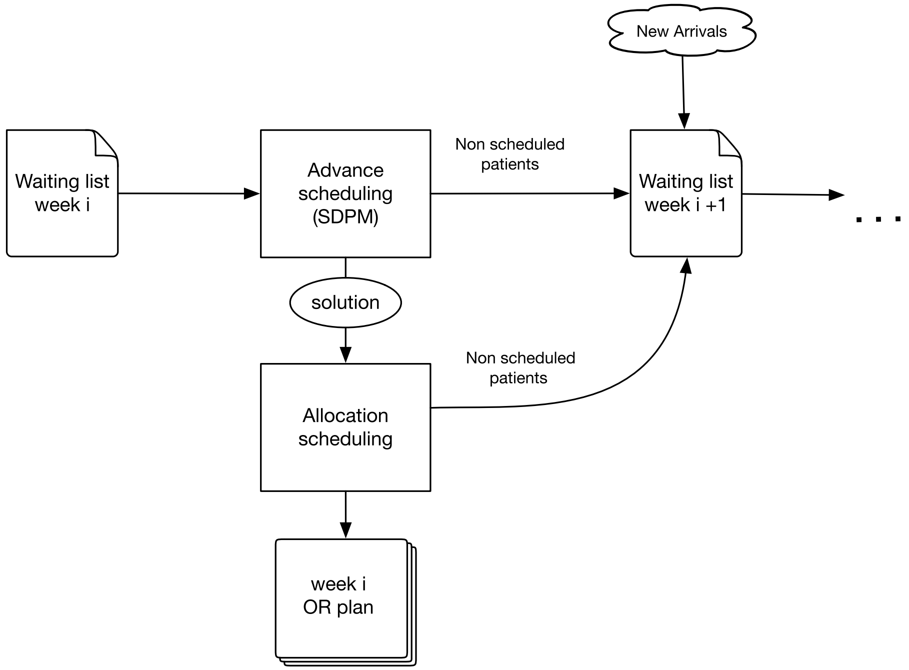

Fig. 1 The two steps of the scheduling procedure

3 Stochastic dynamic programming model

In this section, we formulate a stochastic dynamic programming model (SDPM) to optimally schedule patients and physicians in the hospital ORs. The model solves the *advance scheduling* problem, where patient–physician pairs are allocated on a daily basis for a planning horizon of several weeks. At this level, we determine the daily aggregate number of surgeries to be performed by each physician, considering the capacity of each OR. Next we perform *allocation scheduling*, where surgeries are scheduled into the specific time slots for each OR room (see Fig. 1). In the next section we describe the *allocation scheduling* methodology used.

Within each pathology, patients are first classified according to the number of days they have been waiting to be treated. The maximum admissible waiting time is given by the GES plan; any patient that surpasses this threshold is referred to an alternative hospital at a higher treatment cost. Additionally, the model guarantees that at least a percentage of the OR throughput, defined by the decision maker, is dedicated to non-GES pathologies within the planning horizon. This constraint is incorporated as a border condition, such that a large cost is incurred if it is not satisfied. For this purpose, the model incorporates two additional state variables: the number of GES and non-GES surgeries performed until period t. The model also considers physician-dependent surgery times.

We consider a hospital with J operating rooms that are shared among I pathologies and a planning horizon of T days. In each period of time, we determine the optimal surgery allocation for each pathology i to minimize the total expected cost incurred when referring GES patients to another facility. We use the following notation:

Parameters

- T: length of the planning horizon, in days.
- I: number of pathologies treated, where pathology i has the following parameters:
  - g_i: maximum number of waiting days according to the GES plan (Opportunity Guarantee).
  - λ_{it}: arrival rate of patients on day t.
  - C_i: additional cost if the surgery is transferred to an alternative provider (referral cost). This cost is incurred when a patient reaches the maximum waiting time without receiving treatment.
  - S_{ik1}: number of patients with pathology i at the beginning of the planning horizon ($t = 1$) that has been waiting for k days.
  - G and $\hat{G}$ are the sets of pathologies included and not included in the GES plan, respectively. Thus, we have that $C_i = 0, \forall i \in \hat{G}$.
  - α: minimum fraction of non-GES surgeries required during the planning horizon.
- J: number of operating rooms at the hospital, where OR j is characterized by:
  - ν_{jt}: total hours available on day t for surgeries in operating room j.
  - E_j: set of pathologies that can be treated in operating room j. Additionally, $E_{ij} = 1$ if pathology i can be operated in OR j and 0 otherwise.
- M: total number of physicians, where physician m is characterized by:
  - a^a_{mt}, a^p_{mt}: number of hours that physician m has available on the morning (a^a_{mt}) and the afternoon (a^p_{mt}) of day t. Note that this parameter may change on a weekly basis according to the physician's availability.
  - $p_{mi} = 1$ if physician m operates on pathology type i and 0 otherwise. In the general setting, physicians operate on only some pathologies.
  - H_{im}: average time required by physician m to operate on pathology i.
- L: total number of anesthesiologists, where anesthesiologist l is characterized by:
  - $u^a_{lt}, u^p_{lt} = 1$ if anesthesiologist l is available in the morning (u^a_{lt}) and the afternoon (u^p_{lt}) of day t respectively. Note that this parameter may change on a weekly basis according to the anesthesiologist's availability.
  - N_{il}: average time required by anesthesiologist l for pathology i.
- $H_{ilm} = H_{im} + N_{il}$: average time required by physician m and anesthesiologist l for pathology i.

State variables

- S_{ikt}: number of patients with pathology i waiting for k days at the beginning of day t.
- $\hat{G}_t$: number of non-GES surgeries performed from the beginning of the planning horizon until t.
- G_t: number of GES surgeries performed from the beginning of the planning horizon until t.

Decision variables

- x_{ikjt}: number of patients with pathology i waiting for k days that are operated on in OR j on day t.
- y^a_{imljt}, y^p_{imljt}: number of type i surgeries assigned to physician m with anesthesiologist l on day t in OR j, in the morning (AM) or afternoon (PM), respectively.

- z^a_{ljt}, z^p_{ljt}: 1 if anesthesiologist l is assigned to OR j on day t in the morning (AM) and afternoon (PM), respectively, and 0 otherwise.
- z^a_{jt}, z^p_{jt}: 1 if an anesthesiologist is assigned to OR j on day t in the morning (AM) and afternoon (PM), respectively, and 0 otherwise.

Random variables

- η_{it}: total number of patients with pathology i that arrive during period t, $\mathbb{E}(\eta_{it}) = \lambda_{it}$.

We define $f^*_t(\mathbf{S}_t, \hat{G}_t, G_t)$ as the minimum expected cost from period t until the end of the planning horizon if the state variables are equal to $\mathbf{S}_t$, $\hat{G}_t$, and G_t at the beginning of period t

$$\textbf{SDPM} \qquad f^*_t(\mathbf{S}_t, \hat{G}_t, G_t) = \min_{\mathbf{x}_t, \mathbf{Y}_t} \sum_i C_i S_{ig_{i+1}t} + \mathbb{E}(f^*_{t+1}(\mathbf{S}_{t+1}, \hat{G}_{t+1}, G_{t+1})),$$

where $\mathbb{E}(f^*_{t+1}(\mathbf{S}_{t+1}, \hat{G}_{t+1}, G_{t+1}))$ is the expected optimal cost from period $t+1$ onwards, with respect to $\{\eta_{it}\}_{i=1}^I$; with the following constraints:

1. **Consistency**
   - Definition of auxiliary variable x_{ijt} as the total number of patients with pathology i that are operated on in OR j on day t.

$$x_{ijt} = \sum_{k=1}^{g_i} x_{ijkt}, \quad \forall i, \forall j, \forall t. \tag{1}$$

   - Definition of auxiliary variable y^r_{imjt} as the total number of type i surgeries assigned to physician m in OR j on day t and shift r (a (am) or p (pm)).

$$y^r_{imjt} = \sum_{l=1}^{L} y^r_{imljt}, \quad \forall i, \forall m, \forall j, \forall t, r = \{a, p\}. \tag{2}$$

   - Definition of auxiliary variable z^r_{jt} as 1 if an anesthesiologist is assigned to OR j on day t and shift r (a(am) or p (pm)) and 0 otherwise.

$$z^r_{jt} = \sum_{l=1}^{L} z^r_{ljt}, \quad \forall j, \forall t, r = \{a, p\}. \tag{3}$$

2. **Flow and assignment**
   - Update of waiting list at the beginning of day $t+1$.

$$S_{ik+1t+1} = S_{ikt} - \sum_{j=1}^{J} x_{ikjt}, \quad \forall i, k \in [1, g_i + 1], \tag{4}$$

$$S_{i1t+1} = \eta_{it}, \quad \forall i, \tag{5}$$

$$S_{ik1} = \text{waiting list on day 1}, \quad \forall i \text{ and } 1 \le k \le g_i + 1. \tag{6}$$

   - Update of number of non-GES surgeries performed until time $t+1$.

$$\hat{G}_{t+1} = \hat{G}_t + \sum_{i \in \hat{G}} \sum_j \sum_k x_{ikjt}. \tag{7}$$

- Update of number of GES surgeries performed until time $t+1$.

$$G_{t+1} = G_t + \sum_{i \in G} \sum_j \sum_k x_{ikjt}. \tag{8}$$

- For each pathology, the number of scheduled surgeries must be equal to the number of surgeries assigned to the physicians for each day t.

$$x_{ijt} = \sum_{m=1}^{M} \left(y^a_{imjt} + y^p_{imjt} \right), \quad \forall i, \forall j, \forall t. \tag{9}$$

- Operating rooms are equipped to serve only certain types of pathologies.

$$x_{ikjt} = 0, \quad \forall i \notin E_j, \forall k, \forall t, \tag{10}$$

or equivalently,

$$\sum_{k=1}^{g_i} x_{ikjt} \le \mathcal{M} E_{ij}, \quad \forall i, \forall j, \forall t, \tag{11}$$

where $\mathcal{M}$ is a large number.

3. **Capacity/availability**

- The number of surgeries must be less than or equal to the number of patients with the corresponding pathology on each day t.

$$\sum_{j=1}^{J} x_{ijt} \le \sum_{k=1}^{g_i} S_{ikt} \quad \forall i, \forall t. \tag{12}$$

- Physicians operate on only a subset of pathologies.

$$\sum_{j=1}^{J} \left(y^a_{imjt} + y^p_{imjt} \right) \le \mathcal{M} p_{mi}, \quad \forall m, \forall i, \forall t, \tag{13}$$

where $\mathcal{M}$ is a large number.

- The number of surgery hours assigned to each physician must be less than or equal to his/her availability.

$$\sum_{j=1}^{J} \sum_{i=1}^{I} \sum_{l=1}^{L} H_{iml} y^r_{imljt} \le a^r_{mt}, \quad \forall m, \forall t, r = \{a, p\}. \tag{14}$$

- There is no more than one anesthesiologist per OR for any given day and shift.

$$z^r_{jt} \le 1, \quad \forall j, \forall t, r = \{a, p\}. \tag{15}$$

- Anesthesiologist's availability at each shift r and day t.

$$\sum_{j=1}^{J} z^r_{ljt} \le u^r_{lt}, \quad \forall l, \forall t, r = \{a, p\}. \tag{16}$$

- Guarantee that all scheduled surgeries have an anesthesiologist assigned to the OR.

$$\sum_{i=1}^{I}\sum_{m=1}^{M} y^r_{imjt} \leq \mathcal{M} z^r_{jt}, \quad \forall j, \forall t, r = \{a, p\}, \tag{17}$$

$$z^r_{jt} \leq \sum_{i=1}^{I}\sum_{m=1}^{M} y^r_{imjt}, \quad \forall j, \forall t, r = \{a, p\}. \tag{18}$$

- The number of surgeries in an OR cannot exceed its capacity on each day.

$$\sum_{m=1}^{M}\sum_{i=1}^{I}\sum_{l=1}^{L} H_{iml}\left(y^a_{imljt} + y^p_{imljt}\right) \leq \nu_{jt}, \quad \forall j, \forall t. \tag{19}$$

Finally, the border conditions are as follows:

$$f^*_{T+1}(\mathbf{S}_{T+1}, \hat{G}_{T+1}, G_{T+1}) = \begin{cases} \sum_i C_i |S_{iT+1}|, & \text{if } \frac{\hat{G}_{T+1}}{\hat{G}_{T+1}+G_{T+1}} \geq \tilde{\alpha} \\ \mathcal{M} & \text{otherwise.} \end{cases},$$

where $C_i \geq 0$ is the treatment cost of the remaining patients at the end of the planning horizon, $|S_{i,T+1}| = \sum_{k=1}^{g_i} S_{ikT+1}$, and $\tilde{\alpha} = \min\left\{\alpha, \frac{\sum_{i\in\hat{G}}\sum_k S_{ik1}}{\sum_{i\in I}\sum_k S_{ik1}}\right\}$. The inequality $\frac{\hat{G}_{T+1}}{\hat{G}_{T+1}+G_{T+1}} \geq \tilde{\alpha}$ guarantees a minimum fraction of non-GES surgeries of the total number of surgeries within the planning horizon. As discussed previously, in the absence of this constraint, there are no incentives to treat patients with pathologies not included in the GES plan.

The dimensions of the state space for this problem become extremely large and intractable even for relatively small problems. For example, if we consider a total of 5 pathologies, a maximum of 20 waiting days and 20 patients waiting for each, then the state space takes 20^{100} possible values considering only the waiting list. Therefore, in Sect. 4, we develop an heuristic, based on a deterministic version of SDPM, to find good approximations of the optimal solution for the stochastic model. The following proposition provides a lower bound for the total expected cost obtained by the stochastic model, which we later use to measure the quality of these approximations.

Proposition 1 *The expected value of the objective function of the model below* ($\mathbb{E}_{\boldsymbol{\eta}}[LB(\boldsymbol{\eta})]$) *provides a lower bound for the SDPM problem.*

$$LB(\boldsymbol{\eta}) = \min_{\mathbf{x},\mathbf{y},\mathbf{S}} \sum_{i=1}^{I}\sum_{t=1}^{T} C_i S_{ig_i+1t}, \tag{20}$$

subject to: (1)–(6), (9)*, and* (11)–(19) *and the following constraint:*

- *The total number of GES-related surgeries must maintain the defined ratio:*

$$WA_w + \frac{\sum_{i\in\hat{G}}\sum_{j=1}^{J}\sum_{t=1}^{T} H_{im}\left(y^a_{imjt} + y^p_{imjt}\right)}{\sum_{i\in\{\hat{G}\cup G\}}\sum_{j=1}^{J}\sum_{t=1}^{T} H_{im}\left(y^a_{imjt} + y^p_{imjt}\right)} \geq \tilde{\alpha}\left(W + \frac{T}{7}\right), \tag{21}$$

where A_w is the average number of hours of non-GES surgeries in the last W weeks.

Proof: It follows directly since any solution of the **SDPM** problem is feasible for this problem. We notice that (20) solves for the optimal schedule with perfect information regarding future patients' arrivals, and therefore, its expected value outperforms the solution of the stochastic model that considers all possible random arrivals when making scheduling decisions. We refer to this model as the perfect information model (**PIM**). □

4 The GES-PROG algorithm

In this section we describe the heuristic developed to find good approximations for the optimal solution of the **SDPM**. The heuristic runs on a rolling horizon basis, i.e., the waiting list is updated at the beginning of each week and the model is solved for the complete planning horizon. The solution for only the current week is then implemented.

The current practice at the *Instituto de Neurocirugía* consists of weekly meetings with the OR manager, a physician, three interns, and a nurse in charge of scheduling. Based on the waiting list, and without explicitly considering future (uncertain) arrivals, the committee schedules surgeries for the coming week. This procedure is highly manual and relies heavily on the personal experience of the committee members. Priority is given to patients with deadlines that expire during the immediate week. The heuristic proposed in this section, namely GES–PROG, is similar to the current decision process at the hospital in terms of considering only the waiting list and not future stochastic arrivals when making scheduling decisions. However, we note that our heuristic found an optimal solution for the optimization model, which outperforms the personal experience of the decision-making team at the hospital. The objective function in this model is the minimization of the total GES referral costs. However, to maximize the utilization of the operating rooms, we include a reward for programming additional pathologies even if they do not account for referral costs.

After obtaining the *advance scheduling* solution for the current week, we solve the *allocation scheduling* problem to obtain the final daily schedule. The *allocation scheduling* problem solved and the methodology used are described in Azar et al. (2017). In Azar et al. (2017), we develop an efficient time-indexed scheduling formulation that is able to compute, an optimal solution for the allocation problem. The solution to that problem is computed in a few minutes thanks to the fact that we reduce the number of patient/physician pairs through the GES-PROG algorithm. In order to schedule as many surgeries as possible, we maximize the throughput of surgeries.

Because the *advance scheduling* model considers the aggregate daily availability of physicians, there may be infeasibilities when scheduling daily surgeries. There could be cases where the only feasible solution is to schedule a physician in two ORs at overlapping times. The *allocation scheduling* model finds the solution that maximizes the throughput. All surgeries that cannot be scheduled are returned to the waiting list for future scheduling.

Algorithm GES-PROG

1: **Initialization**: S_{ik1} = number of patients with pathology i that have been waiting for k days at the beginning of the planning horizon.

2: **Solve Advance Scheduling:**

$$LB1 = \min_{\mathbf{x},\mathbf{y},\mathbf{S}} \sum_{i=1}^{I}\sum_{t=1}^{T} C_i S_{ig_i+1t} - 10^{-3}\sum_{i=1}^{I}\sum_{j=1}^{J}\sum_{k=1}^{g_i}\sum_{t=1}^{T} k\, x_{ijkt},$$

subject to (1)–(19), (21), and $S_{i1t} = 0, \forall t$.

3: **Solve the *allocation scheduling***: Using the weekly assignment from Step 2, we use the operational model described in Azar et al. (2017) to determine which time slot, physician, and OR to assign to each patient. If some surgeries cannot be scheduled due to infeasibility, they are returned to the waiting list.

4: **Implement the solution for the first week**: Number of all surgeries scheduled by the model, for GES and non-GES procedures, by physician, operating room, waiting time and day: $\mathbf{x} = x_{ikjt}, \forall i, \forall k, \forall j, t = \{1, \ldots, 7\}$.

We also studied a second heuristic, which accounted for the expected number of future arrivals on the weekly schedule. The purpose of this heuristic was to avoid myopic decisions when using just the current waiting list without considering that patients with expensive and short deadline surgeries might arrive in future weeks, which might "displace" already scheduled patients. The computational analysis, described in Sect. 6, showed that this heuristic, systematically, underperformed GES–PROG described above, and therefore, we only report the performance of the latter.

5 Data

The *Instituto de Neurocirugía* has five ORs: one for emergency procedures and four for elective surgeries. Preparation and recuperation from anesthesia are performed on site. One of these four OR units is used four days a week for pediatric and one day a week for adult surgeries. Log data from 2015 and 2016, consisting of diagnosis, surgeon and anesthesiologist in charge, and surgery and anesthetic times, were analyzed. Among all types of surgeries performed at the hospital, 19 accounted for 80% of all procedures. Two types of surgery performed at the hospital are classified as GES and therefore are subject to the opportunity guarantee.

Brain tumor surgeries are divided into two diagnoses depending on severity; thus, they typically last either 4 or 7 h. This information is known prior to the scheduling process. Additionally, we obtained the average number of surgeries performed at the hospital per day. Given that the defection rate is low, these estimates are used as a good approximation of the average arrival rates of new patients; see Appendix A, Table 5. We also calculated the average surgery time for each diagnosis and surgeon; see Table 7. We note that, for a given diagnosis, the average surgery time varies significantly among surgeons. For example, for HNP, the range of surgery time among physicians is $[1, 3]$ hours with an average of 1.4 h. Finally, Table 6 shows the opportunity guarantees and costs for GES diagnoses, for which, according to the hospital's OR Manager, the referral costs are double the in-hospital costs.

Based on this data analysis, we generated a waiting list of patients equivalent to 100 days. According to the OR Manager, this scenario represents a normal traffic case. For the simulations, the waiting list is updated weekly on a rolling horizon basis by eliminating those patients that have received treatment and incorporating new randomly generated arrivals. The initial waiting list includes 37 GES patients and 139 non-GES patients. A complete description of the waiting list is given in Appendix A.2.

6 Simulation-based performance analysis

In this section we present a computational study of the GES-PROG algorithm. We divide this study into two parts. First, in Sect. 6.1, we study the relationships between the cost of referrals and the due dates of the GES pathologies, and how they affect the solutions given by GES-PROG. Next, in Sect. 6.2 we study larger instances (more pathologies and larger waiting lists) to understand the effect of the system's size and load on the performance of our algorithm.

The purpose of this analysis is to study the performance of GES-PROG under different settings, and provide a sensitivity analysis for the relevant parameters. Later, in Sect. 7, we validate its usefulness in a real case scenario, by comparing the results of our algorithm against the solutions proposed by the OR manager.

All computational experiments in Sects. 6 and 7 were done in an Intel Xeon CPU E5-2640 v2 2.00 GHz with 24 GB of RAM running CentOS 7 as operating system. All algorithms were implemented in Python 3.5, using Gurobi 7.5 as solver for the corresponding optimization model. In each analysis, we simulated a single operating room with a daily availability of 8 h, and studied the output of the GES-PROG algorithm by simulating 10 years of 30 weeks each. We computed weekly average metrics. The size of the instances was chosen in order to obtain a coefficient of variation of less than 0.001 for the estimation of all metrics.

6.1 GES-PROG sensitivity analysis

In this subsection we present the sensitivity analysis of the GES-PROG algorithm, with respect to due dates and referral costs for GES pathologies. For this purpose, we analyze the results of the algorithm for a set of test cases with two GES pathologies, A and B. We consider that the surgery times for pathologies A and B are 1 and 2 h respectively, independently of the physician ($H_{Am} = 1, \forall m$ and $H_{Bm} = 2, \forall m$). We fixed the due date of pathology A to 20 days and the referral cost to 1 ($g_A = 20$ and $C_A = 1$). In what follows, we analyze the effect of the due date and referral cost of pathology B (g_B and C_B) on the performance indicators of the algorithm, such as the number of referrals and average patients' waiting time. In each run, we start with an empty waiting list and study the output for different arrival rates of new patients. Each pair of arrival rates for surgeries A and B, was selected in order to keep the combined rate constant and equivalent to 10% higher than the available OR capacity, i.e. a combined arrival rate equal to 44 h of surgery per week. For all the experiments in this subsection, the worst coefficient of variation was 0.00076. Figures 2 and 3 summarize the results of this analysis, showing the average weekly referrals for each pathology.

Figure 2 shows the effect on referrals when varying the due date of pathology B in $g_B \in \{5, 10, 20, 30, 40\}$ for each arrival rate pair, while keeping the referral cost of B constant and equal to 1 ($C_B = 1$). Figures 2a, b show the referrals of A and B respectively. We observe that the increase in the due date of B, allows the algorithm to accommodate more surgeries without requiring referrals, due to the added flexibility. Figure 2b shows that the increase in the due date of pathology B reduces the number of patients that need to be referred, while avoiding referrals of pathology A, as observed in Fig. 2a. As the arrival rate of pathology B decreases to 0, the referrals of A increase, but only when the arrival rate of A increases significantly, since the length of this pathology is only half of B.

Figure 3 shows the number of referrals as a function of the referral cost of B, for $C_B \in \{0.1, 0.5, 1, 2, 5, 10\}$ for each arrival rate pair. For these computational experiments, we set the due dates of both pathologies in 20 days. Figure 3a, b show the referrals of A and B respectively. We observe that the algorithm has a binary behavior: when referral cost per hour of B is less than the referral cost per hour of A, then the algorithm only refers A-type surgeries, whereas only B-type surgeries are referred when the cost per hour of B is lower.

It is important to note that, since the instances were small enough, in all these experiments we were able to compute the lower bound described in Proposition 1. In all cases the algorithm's solution had the same cost as the lower bound, .i.e. GES–PROG led to the optimal solution.

In what follows we study the effect of due dates on the average patient's waiting time. For this purpose, we analyze a set of instances with the two GES pathologies currently treated at the Instituto de Neurocirugía, using the actual values of the referrals costs and due dates

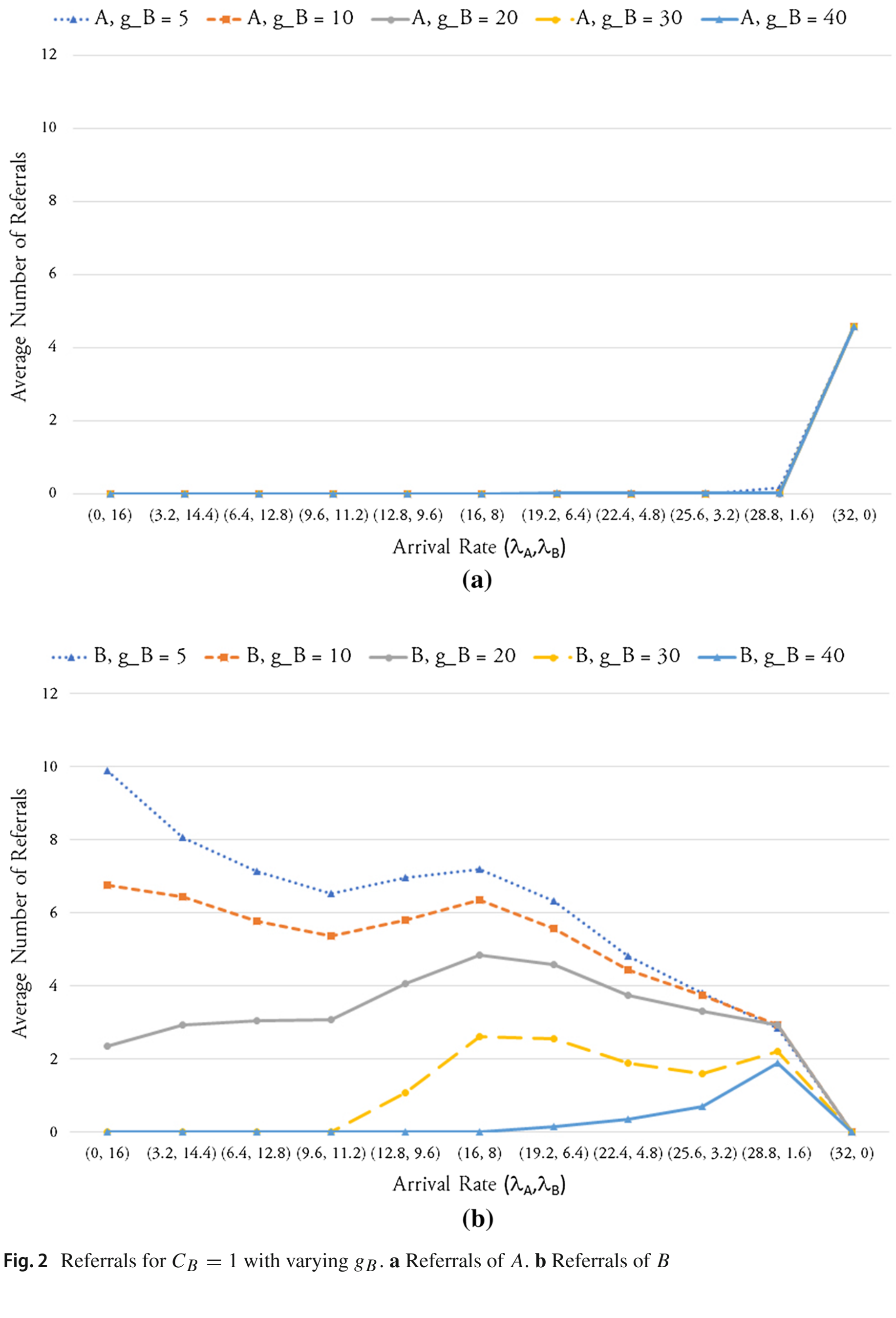

Fig. 2 Referrals for $C_B = 1$ with varying g_B. **a** Referrals of A. **b** Referrals of B

(see Table 6), and assuming all physicians can operate all pathologies. For each case, we compute the average values for several important performance metrics while scaling the due dates by different factors. The purpose of this study is to determine the effects on the patients' waiting times, if, for example, the health authority decides to modify the current due dates

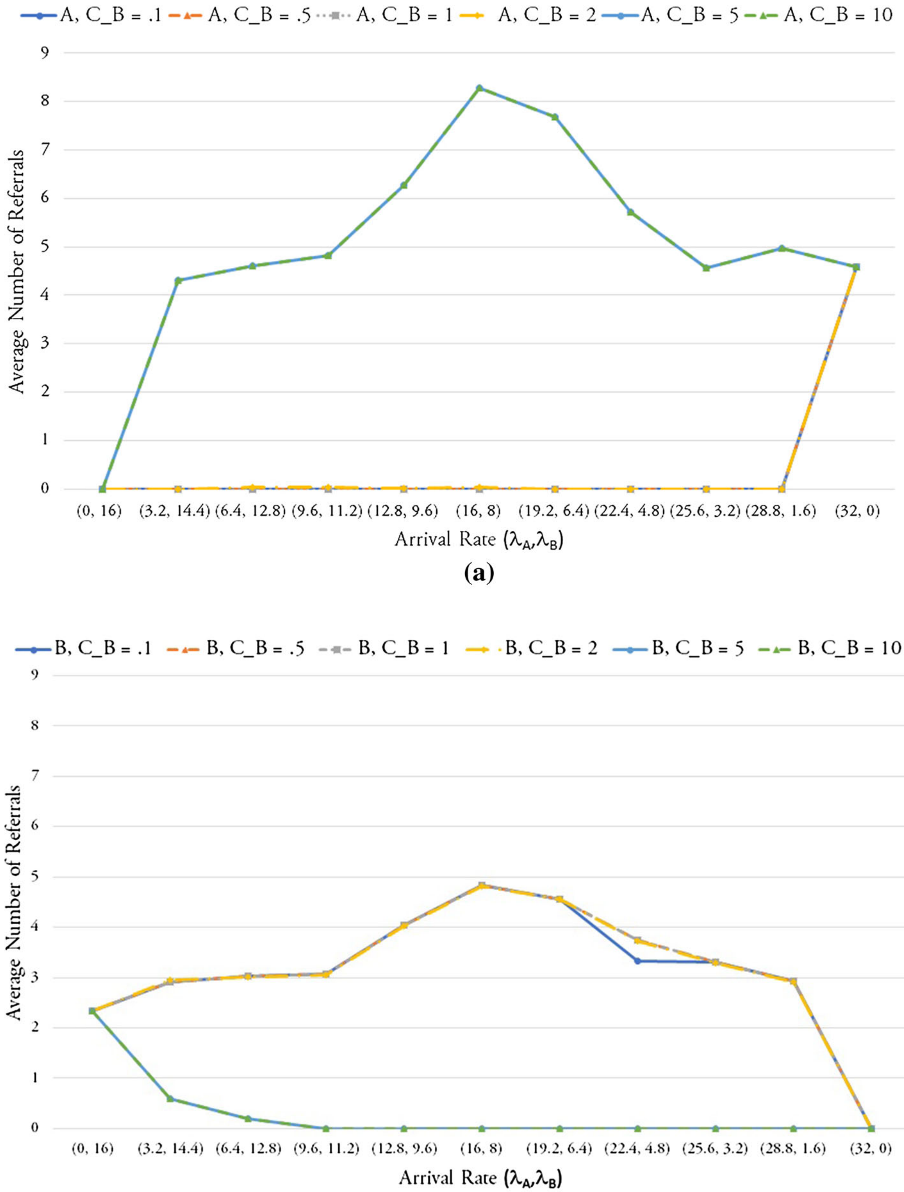

Fig. 3 Referrals for $g_B = 20$ with varying C_B. **a** Referrals of A. **b** Referrals of B

of the GES pathologies. For these instances, we also computed the lower bound described in Proposition 1, showing that there is no difference in the optimal value between that solution and the one obtained by GES-PROG. Details of the performance metrics computed for these experiments can be found in Appendix B.1.

Table 1 Results for scaling g_A

Metric/scaling factor	0.1	0.5	0.8	1.0	1.3	1.5	2.0
Average GES A patient waiting time (days)	2.6	3.1	3.2	3.2	3.2	3.2	3.2
Average GES B patient waiting time (days)	3.4	3.4	3.4	3.4	3.4	3.4	3.4

Table 2 Results for scaling g_B

Metric/scaling factor	0.1	0.5	0.8	1.0	1.3	1.5	2.0
Average GES A patient waiting time (days)	3.5	3.2	3.2	3.2	3.2	3.2	3.2
Average GES B patient waiting time (days)	1.7	3.2	3.4	3.4	3.4	3.4	3.4

Table 3 Performance analysis for larger instances and different load factors

	Light	Normal	Heavy
OR utilization (%)	0.549	0.934	0.946
Normalized average GES patient waiting time (days)	0.159	0.499	0.785
Average non GES patient waiting time (days)	4.982	13.461	19.452
Average HNP referrals (per week)	0.000	0.000	0.006
Average TE referrals (per week)	0.000	0.006	0.457
Average number of served patients (per week)	17.730	31.648	33.810
Average number of patients assigned by *advance scheduling* (per week)	18.069	35.051	37.480

Scaling the due dates without changing the referral cost does not affect the patients' waiting time. Tables 1 and 2 show the average waiting time for each type of patient when scaling the due dates for surgery A (g_A) and B (g_B) respectively. In both cases, the average waiting time remains constant for all scaling factors, except 0.1, where the due dates are shorter than a week. In this case, the algorithm cannot cope with all the arrivals and must refer several patients, modifying the average waiting time.

6.2 Impact of patients' load on GES-PROG's performance

To evaluate the effect of the patients' load in the system on the performance of GES-PROG , we stressed the algorithm under different load levels. For this purpose, we analyze the performance of GES-PROG considering 28 different pathologies treated at Instituto de Neurocirugía, including the two GES pathologies. The initial patient waiting list was generated for 8 weeks of arrivals, and the largest coefficient of variation in all the experiments was 0.0068. Several performance metrics were computed for different load levels. The average results are summarized in Table 3.

The *Normal Load* setting was simulated by using the current arrival rates for different types of patients at the Instituto de Neurocirugía. The *Light Load* and *Heavy Load* settings were simulated by scaling the normal arrival rates by 0.5 and 2, respectively. Increasing the arrival rates barely changes the average number of referrals because the algorithm delays non-GES surgeries. This delay substantially increases the average waiting time of non-GES patients, as well as that of GES patients, who now have to wait almost 80% of the total due date before their surgery is performed. Another important result from these experiments is that the number of patients assigned by *advance* and *allocation scheduling* (see Fig. 1 in Sect. 3) are similar. Therefore, the aggregated daily physicians' time allocation that takes place at the *advance scheduling* does not substantially affect the algorithm's performance. In fact, in all settings, more than 90% of the patients assigned by *advance scheduling* are actually scheduled in the *allocation scheduling*.

Finally, as in the previous experiments, the we also computed the lower bound described in Proposition 1 for comparison. In the three scenarios, the lower bound was able to schedule all GES patients without requiring referrals. For GES-PROG , this was only true for the light and normal loads. In the heavy load setting, the GES-PROG had few referrals per week (0.46). This number is small considering that the setting has a large number of patients arriving and that the computation of the lower bound described in Proposition 1 (which is obtained by solving an optimization problem) takes several orders of magnitude more time to solve.

7 Performance validation: application at the Instituto de Neurocirugía

In order to validate the simulation based analysis, and further understand the performance of GES-PROG, in this section we compare our algorithm against the current practice at the *Instituto de Neurocirugía* in a real case setting. For this purpose, we developed a simulation tool in which the hospital's OR Manager performs weekly scheduling for a given initial waiting list with stochastic patient arrivals. We also used our heuristic to solve the problem for the same simulation instances.

Method the simulation process is as follows. The hospital's OR Manager scheduled surgeries on a weekly basis for a total of 15 weeks or 75 days. This planning horizon was chosen by the OR Manager given his time availability to perform the study. We remark that in real life, the weekly schedule requires approximately 2 h of six professionals, and therefore, this simulation process took approximately 30 h for the neurosurgeon OR manager.

Four OR rooms were considered, with one functioning only once a week (the rest of the time is dedicated to pediatric surgery). The regular OR hours are from 8 am to 5 pm. The initial patients waiting list was generated considering 12 weeks of arrivals. As suggested by the OR Manager, we assumed that all physicians were available during the considered planning horizon. We note that the hospital waiting list consists of patient–physician pairs, i.e., surgeons are assigned to patients; therefore, there is one less degree of flexibility when determining the weekly schedule. During the simulation process, we observed the following behavior of the hospital OR Manager: he first scheduled all GES pathologies whose opportunity guarantee would expire during the current week and then scheduled non-GES patients with the longest waiting times. Finally, if OR capacity remained, he balanced the physicians' workloads. We also observed that the OR Manager utilized discretionary overtime.

The scheduling results from the hospital's OR Manager simulations were compared with those obtained using GES-PROG. In the latter case, we analyzed two cases for the patient–physician pairs: (i) preassigned, as considered by the hospital's OR Manager, and (ii) determined by the scheduling model.

Results For the case where the hospital's OR Manager utilized overtime, the scheduling led to an average OR utilization of 87.9% with 30.5 surgeries per week and approximately 30% corresponding to GES surgeries. The average overtime was 58.8 minutes, and each OR had overtime scheduled on 21.3% of the days. Non-GES patients that were attended waited an average of 39.4 days, and GES patients waited an average of 24.7 days (29.1 and 19.0

days for herniated nucleus pulposus (HNP) and brain tumor (TU–ENC), respectively). The same analysis was performed after eliminating surgeries that were scheduled past the regular hours of operation. In this case, the OR utilization and average number of weekly surgeries decreased to 82% and 27, respectively.

By contrast, using GES-PROG without allowing overtime, we obtained a significant increase in OR capacity utilization, reaching an average of 91.5%. It is important to note that this increase is achieved without incurring overtime.

Finally, we ran the heuristic without the patient–physician constraint to study the impact on the additional flexibility of determining which physicians will perform which surgeries during the weekly scheduling. In this case, the average number of surgeries increased from 30.5 to 35.2, with an important reduction in number of waiting days from 39.4 to 26.3 for non-GES patients and from 24.7 to 15.7 for GES patients. It is interesting to note that OR utilization in this case decreased slightly from 87.9 to 87.3%, mainly due to the reduction of almost half of the patients on the waiting list. There were fewer patients to schedule; therefore, some slots were left open.

Table 4 shows the detailed results of the performance validation analysis. Tables 11 and 12 in Appendix B.2 present the performance indicators of the weekly simulations.

Table 4 OR Manager versus GES-PROG performance

Performance measures	OR Manager		GES-PROG	
	With overtime	W/o overtime	With assigned patient/physician	W/o assigned patient/physician
OR utilization (SD)	87.9% (0.9)	82.0% (0.8)	91.5% (4.1)	87.3 (1.0)
Average number of surgeries/week (SD)	30.5 (2.1)	27.0 (2.0)	31.4 (4.5)	35.2 (5.1)
GES surgeries/week as fraction of total surgeries (SD)	30.3% (9.3)	31.3% (11.2)	32.5% (9.8)	28.6 (9.1)
Average minutes of overtime, if there is overtime (SD)	58.8 (37.2)	0 (0)	0 (0)	0 (0)
Percentage of days–OR with overtime (SD)	21.3% (41)	0% (0)	0% (0)	0% (0)
Average waiting days for non-GES surgeries (SD)	39.4 (3.4)	39.0 (3.7)	35.4 (4.7)	26.3 (5.2)
Average waiting days for GES surgeries (SD)	24.7 (1.1)	23.2 (1.5)	22.4 (1.6)	15.7 (2.6)
Average waiting days for GES HNP (SD) opportunity guarantee = 45 days	29.1 (2.1)	28.6 (2.1)	22.0 (3.0)	16.5 (3.4)
Average waiting days for TU-ENC (SD) opportunity guarantee= 25 days	19.0 (1.5)	19.1 (1.5)	22.2 (0.3)	15.0 (1.6)
Average length of waiting list for non-GES patients (SD)	71.2 (5.0)	73.8 (4.8)	57.7 (4.6)	35.4 (6.7)
Average length of waiting list for GES patients (SD)	26.4 (1.3)	27.1 (1.3)	20.9 (1.6)	12.0 (2.5)

SD standard deviation

8 Conclusions

In this paper we have developed an SDP model to compute a near-optimal solution for the OR for the *Advanced Scheduling* problem, in reasonable times to be used as a regular tool by a health provider. The problem addresses the issue of patients waiting to be treated, according to the Hospital de Neurocirugía Dr. A. Asenjo's physical and organizational conditions. An important novelty of our model, is the consideration that patients with a subset of diagnoses defined by the GES plan are referred to the private system when the due date opportunity guarantee is not satisfied, resulting in additional expenses for the public system. This cost is model according to a step function that it is different from the cost functions reported in the literature. We consider a planning horizon of several weeks to take into account the complete waiting list with the corresponding due dates of all patients and future arrivals. Due to the extremely large size of the state space of the model, we developed the GES-PROG algorithm proposed in Sect. 4 to find good approximations to the optimal solution.

The proposed algorithm was studied first in a simplified scenario with only two pathologies, and later in a more realistic simulation game with the collaboration of the OR manager at Instituto de Neurocirugía. For this purpose, we collected data from 2015 and 2016 to estimate all the relevant parameters of the model. In the simplified situation, we reported results that show a very good performance even when compared to a lower bound for the objective function for all the demand load scenarios considered: light, medium, and heavy traffic. The reported worst-case scenario has only a small increment in the number of referrals compared to the proposed lower bound.

In the simulation game, the results showed significant improvements in terms of the waiting list size and the OR overtime. It is important to note that the OR manager has a vast experience, with more than 9 years at the hospital, and although his scheduling process does a great job in programming the current week, it does not take into account optimization decisions that might affect future weeks. Due to this, and other discussed factors, our algorithm is able to improve all metrics when compared to the solution decided by the OR Manager. Moreover, flexibility to choose the physician for each patient results in benefits for both the hospital and the patients, and this is something that could only be done within the framework of our model.

From a qualitative point of view, the current allocation process at the hospital takes approximately 1.5–2.0 h per week, where six professionals meet to set the weekly schedule. Applying our heuristics takes only a few seconds and provides an excellent initial allocation for the decision makers. This allows them to use their time more effectively in medical-related issues, and it does not rely on the experience of the main physicians. We also note that this process can be done online, without the physical presence of the actors involved. Due to these benefits, at this stage, we are working closely with the administrators of the hospital to make this algorithm available as a platform connected with their current information system.

One of the important things we were able to note during the development and testing of our algorithm, is that variability in surgery times is an important factor that can lead to large discrepancies between the proposed schedules and the actual realization. Due to this, our future line of research is on how to add this variability within our models and give robust schedules that can control the effect of the variance on the overtime and utilization of ORs.

Acknowledgements This work was partially funded by Anillo Project 1407 and Ingeniería 2030, Corfo.

A Data analysis 2015–2016

A.1 Pathologies characterization

In the following section, we report the analysis performed for the case study in Sect. 7. Tables 5, 6, and 7 show the arrival rates per pathology, surgery times, and GES parameters, respectively.

A.2 Waiting list characterization

Figure 4a, b show the number of patients in each waiting time range at the beginning of the planning horizon for each GES diagnosis. Table 8 contains the number of non-GES patients on the waiting list at the beginning of the planning horizon.

B Experimental results details

B.1 Simulation based performance analysis

Tables 9 and 10 show the results of the simulations conducted to investigate the effect of due date changes. The simulations were performed using real values for C_A, C_B, g_A, and g_B

Table 5 Arrival rates (*GES pathology)

Code	Pathology ID	Arrival rate (patients/day)
ANS	ANEURYSM	0.341
C-EP	EPILEPSY SURGERY	0.060
CRANE	CRANIOPLASTY	0.164
ESC	SCOLIOSIS	0.103
FIJ-C	SPINE FUSION	0.107
HNP	LUMBAR DISC*	0.445
ID	SHUNTING	0.077
LAM	DISCOMPRESSIVE LAMINECTOMY	0.833
R-FIS	CSF FISTULA REPAIR	0.515
STC	CARPAL TUNNEL SYNDROME	0.141
TU-ES	PITUITARY TUMOR	0.097
TU-Q-M	MEDULAR CYST	0.489
TU-B-C	SKULL BASE TUMOR	0.341
TU-ENC	ENCEPHALIC TUMORS*	0.140

Table 6 GES pathology parameters

Code	Cost at hospital	Cost at referral	Opportunity guarantee (days)
HNP	1	2	45
TU-ENC	1	2.9	25

Table 7 Average surgery time per physician (h)

Surgeon	AN	C-EP	CRANE	ESC	FIJ-C	HNP	ID	LAM	R-FIS	STC	TU-ES	TU-Q-M	TU-B-C	TU-ENC
1	4.4	4.5	1.4	2.5	–	1.7	1.1	–	–	–	–	–	–	5.6
2	–	–	1.5	2.9	–	1.6	1.5	2.7	2.4	0.5	–	–	–	3.3
3	4.3	–	–	2.7	2.8	1.9	1.3	2.6	–	–	–	3.0	–	3.6
4	–	–	–	1.4	–	1.1	1.1	1.3	–	0.5	2.2	2.4	–	2.4
5	–	–	–	3.0	–	1.3	–	2.8	–	–	–	–	–	–
6	4.7	–	1.6	2.1	–	1.6	1.9	–	–	0.5	–	–	–	3.7
7	–	–	–	2.3	2.6	1.1	–	1.7	–	–	–	–	–	–
8	–	–	–	–	–	1.6	–	–	–	0.5	3.0	–	–	4.2
9	–	–	–	–	–	–	–	–	–	0.5	–	–	–	–
10	6.2	–	–	–	–	–	–	–	–	–	–	–	6.5	6.3
11	5.4	–	1.1	–	1.3	1.6	1.3	2.0	2.0	0.5	3	–	–	3.8
12	–	–	–	2.7	2.7	1.4	–	–	–	–	–	–	–	–
13	4.4	–	1.6	–	–	–	1.3	–	–	0.6	3	3.1	–	4.4
14	–	–	2.1	–	–	2.1	1.4	–	–	0.5	–	–	–	4.8
15	3.7	–	–	–	–	–	–	–	–	–	–	–	–	–
16	–	–	–	–	–	–	–	–	1.5	–	–	–	–	–
17	–	–	–	–	–	–	–	–	–	0.5	–	–	–	–
18	2.8	–	–	2.4	1.8	1.1	–	1.8	–	0.3	–	3.0	–	2.0
19	–	–	–	1.6	–	1.5	–	–	–	–	2.8	2.7	–	2.9
20	4.5	–	3.1	–	–	1.5	1.0	2.6	–	–	2.9	2.1	5.5	4.7
21	–	–	–	–	–	3.0	–	–	–	–	3.1	4.5	–	3.8

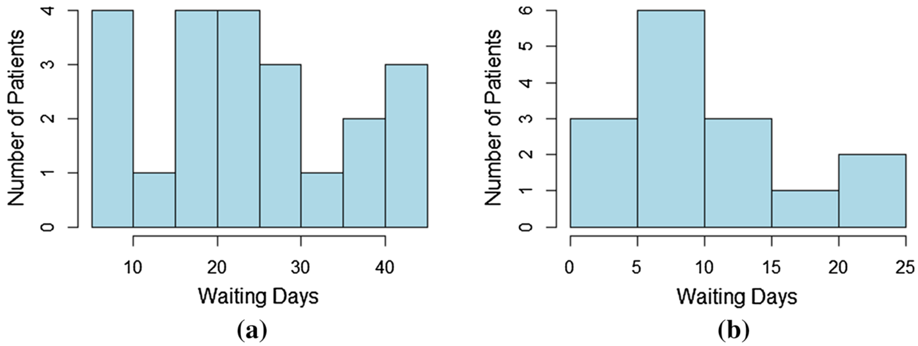

Fig. 4 Waiting list characterization. **a** Waiting time histogram in days for HNP. **b** Waiting time histogram in days for TU-ENC

Table 8 Number of non-GES patients on the waiting list

Diagnosis code	Number of non-GES patients on the initial waiting list
AN	20
C-EP	2
CRANE	9
ESC	29
FIJ-C	5
ID	22
LAM	4
R-FIS	9
STC	21
TU-ES	11
TU-Q-M	4
TU-B-C	3

and assuming that all physicians can operate on all pathologies. For each case, we simulated 10 years of 30 weeks each and computed the average values for each important metric while scaling g_A in Table 9 and scaling g_B in Table 10.

B.2 Performance validation

Tables 11 and 12 show additional details of the case study conducted at Instituto de Neurocirugía to validate the simulation results.

Table 9 Results for changing g_A

Metric/scaling factor	0.1	0.5	0.8	1	1.3	1.5	2
OR utilization (%)	0.093	0.096	0.096	0.096	0.096	0.096	0.096
Average GES patient waiting time (normalized)	0.368	0.138	0.112	0.103	0.094	0.090	0.084
Average GES A patient waiting time (days)	2.621	3.117	3.186	3.186	3.159	3.162	3.186
Average GES B patient waiting time (days)	3.433	3.418	3.418	3.418	3.418	3.418	3.418
Average GES referrals	0	0	0	0	0	0	0
Average number of patients served	9.58	10	10	10	10	10	10
Average number of patients assigned by the *advance schedule*	9.58	10	10	10	10	10	10
Gap (GES-PROG/lower bound)	0	0	0	0	0	0	0

Table 10 Results for changing g_B

Metric/scaling factor	0.1	0.5	0.8	1	1.3	1.5	2
OR utilization (%)	0.065	0.094	0.096	0.096	0.096	0.096	0.096
Average GES patient waiting time (normalized)	0.371	0.159	0.119	0.103	0.087	0.080	0.070
Average GES A patient waiting time (days)	3.530	3.159	3.184	3.186	3.186	3.186	3.186
Average GES B patient waiting time (days)	1.713	3.168	3.405	3.418	3.367	3.373	3.418
Average GES referrals	0	0	0	0	0	0	0
Average number of patients served	7.503	9.813	9.97	10	10	10	10
Average number of patients assigned by the *advance schedule*	7.503	9.813	9.97	10	10	10	10
Gap (GES-PROG/lower bound)	0	0	0	0	0	0	0

Table 11 OR Manager performance with assigned patient–physician pairs

Week	Total number of surgeries	% GES (h) surgeries (%)	Referrals	Overtime (h)	OR capacity utilization (%)
1	33	31	0	3.5	87
2	27	41	0	1.5	82
3	31	39	0	1.4	89
4	35	27	0	4.3	91
5	32	18	0	3.7	88
6	29	13	0	1.4	85
7	29	25	0	1.1	85
8	31	44	0	2.5	91
9	29	25	0	1.6	91
10	31	38	0	0.6	86
11	28	44	0	1.3	89
12	30	37	0	2.9	81
13	29	25	0	2.0	91
14	33	30	0	2.0	91
15	30	43	0	0.7	91

Table 12 GES-PROG performance with assigned patient–physician pairs

Week	Total number of surgeries	% GES surgeries (%)	Referrals	Overtime (h)	OR capacity utilization (%)
1	43	22	0	0	90
2	33	26	0	0	88
3	32	45	0	0	94
4	32	31	0	0	92
5	31	31	0	0	96
6	28	10	0	0	88
7	26	22	0	0	92
8	39	28	0	0	96
9	30	45	0	0	94
10	29	39	0	0	89
11	32	32	0	0	95
12	32	43	0	0	95
13	27	36	0	0	85
14	30	37	0	0	84
15	27	39	0	0	95

References

Addis, B., Carello, G., Grosso, A., & Tànfani, E. (2016). Operating room scheduling and rescheduling: A rolling horizon approach. *Flexible Services and Manufacturing Journal*, *28*(1–2), 206–232.

Aringhieri, R., Landa, P., Soriano, P., Tnfani, E., & Testi, A. (2015). A two level metaheuristic for the operating room scheduling and assignment problem. *Computers & Operations Research*, *54*, 21–34. https://doi.org/10.1016/j.cor.2014.08.014.

Astaraky, D., & Patrick, J. (2015). A simulation based approximate dynamic programming approach to multi-class, multi-resource surgical scheduling. *European Journal of Operational Research*, *245*(1), 309–319.

Azar, M., Barrera, J., Carrasco, R. A., & Mondschein, S. (2017). Operating room scheduling with variable procedure times. In *Proceedings of the 13th workshop on models and algorithms for planning and scheduling problems*. Seeon Abbey, Germany.

Beliën, J., & Demeulemeester, E. (2007). Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research*, *176*(2), 1185–1204.

Brandeau, M. L., Sainfort, F., & Pierskalla, W. P. (2004). *Operations research and health care: A handbook of methods and applications* (Vol. 70). Berlin: Springer.

Bruni, M., Beraldi, P., & Conforti, D. (2015). A stochastic programming approach for operating theatre scheduling under uncertainty. *IMA Journal of Management Mathematics*, *26*(1), 99–119.

Cardoen, B., Demeulemeester, E., & Beliën, J. (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research*, *201*(3), 921–932.

Catalina De Améstica, R. (2014). Informe revela que un 56% de los pacientes "no Auge" lleva más de un año esperando una cirugía. *Diario La Segunda*. http://www.lasegunda.com/movil/detallenoticia.aspx?idnoticia=934670. Accessed 2 May 2018.

Denton, B., Viapiano, J., & Vogl, A. (2007). Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, *10*(1), 13–24.

Dexter, F., Dexter, E. U., Masursky, D., & Nussmeier, N. A. (2008). Systematic review of general thoracic surgery articles to identify predictors of operating room case durations. *Anesthesia & Analgesia*, *106*(4), 1232–1241.

Duma, D., & Aringhieri, R. (2015). An online optimization approach for the real time management of operating rooms. *Operations Research for Health Care*, *7*, 40–51.

Gomes, C., Almada-Lobo, B., Borges, J., & Soares, C. (2012). Integrating data mining and optimization techniques on surgery scheduling. In S. Zhou, S. Zhang, & G. Karypis (Eds.), *ADMA* (pp. 589–602). Berlin: Springer.

Guerriero, F., & Guido, R. (2011). Operational research in the management of the operating theatre: A survey. *Health Care Management Science*, *14*(1), 89–114.
Hans, E., Wullink, G., Van Houdenhoven, M., & Kazemier, G. (2008). Robust surgery loading. *European Journal of Operational Research*, *185*(3), 1038–1050.
Jebali, A., Alouane, H. A., & Ladet, P. (2006). Operating rooms scheduling. *International Journal of Production Economics*, *99*(1–2), 52–62.
Landa, P., Aringhieri, R., Soriano, P., Tànfani, E., & Testi, A. (2016). A hybrid optimization algorithm for surgeries scheduling. *Operations Research for Health Care*, *8*, 103–114.
Lenz, R. (2007). Proceso político de la reforma auge de salud en Chile: algunas lecciones para América Latina: una mirada desde la economía política. *CiEPLAN Santiago de Chile*.
Magerlein, J. M., & Martin, J. B. (1978). Surgical demand scheduling: A review. *Health Services Research*, *13*(4), 418.
Min, D., & Yih, Y. (2010). An elective surgery scheduling problem considering patient priority. *Computers & Operations Research*, *37*(6), 1091–1099.
Molina, J. M., & Framinan, J. M. (2009). Testing planning policies for solving the elective case scheduling phase: A real application. In *Proceedings of the 35th international conference on operational research applied to health services (ORAHS)*. Leuven, Belgium
Patrick, J., & Puterman, M. L. (2008). Reducing wait times through operations research: Optimizing the use of surge capacity. *Healthcare Policy*, *3*(3), 75.
Pinedo, M. L. (2012). *Scheduling: Theory, algorithms, and systems* (4th ed.). Berlin: Springer.
Samudra, M., Van Riet, C., Demeulemeester, E., Cardoen, B., Vansteenkiste, N., & Rademakers, F. E. (2016). Scheduling operating rooms: Achievements, challenges and pitfalls. *Journal of Scheduling*, *19*(5), 493–525.
Santibáñez, 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*(3), 269–282.
Stepaniak, P. S., Heij, C., & De Vries, G. (2010). Modeling and prediction of surgical procedure times. *Statistica Neerlandica*, *64*(1), 1–18.
Tànfani, E., & Testi, A. (2010). A pre-assignment heuristic algorithm for the master surgical schedule problem (MSSP). *Annals of Operations Research*, *178*, 105–119.
Testi, A., & Tanfani, E. (2009). Tactical and operational decisions for operating room planning: Efficiency and welfare implications. *Health Care Management Science*, *12*, 363–373.
Vansteenkiste, N., Lamote, C., Vandersmissen, J., Luysmans, P., Monnens, P., De Voldere, G., et al. (2012). Reallocation of operating room capacity using the due-time model. *Medical Care*, *50*(9), 779–784.
World Health Organization. (2017). *Surgical care systems strengthening: Developing national surgical, obstetric and anaesthesia plans*. Geneva: World Health Organization.
Zúñiga-Fajuri, A. (2007). Sistemas sanitarios y reforma AUGE en Chile. *Acta Bioethica*, *13*(2), 237–245.

Affiliations

Javiera Barrera[1] · Rodrigo A. Carrasco[1] · Susana Mondschein[1] · Gianpiero Canessa[1] · David Rojas-Zalazar[2]

Javiera Barrera
javiera.barrera@uai.cl

Rodrigo A. Carrasco
rax@uai.cl
http://www.rax.cl

Gianpiero Canessa
gianpiero.canessa@edu.uai.cl

David Rojas-Zalazar
darojas@med.uchile.cl

[1] Faculty of Engineering and Sciences, Universidad Adolfo Ibáñez, Santiago, Chile

[2] Institute of Neurosurgery Dr. A. Asenjo, Universidad de Chile, Santiago, Chile