Scheduling operating rooms with consideration of all resources, post anesthesia beds and emergency surgeries

Guillermo Latorre-Núñez a, Armin Lüer-Villagra a,d, Vladimir Marianov a,* Carlos Obreque b, Francisco Ramis b, Liliana Neriz c

a Department of Electrical Engineering, Pontificia Universidad Católica de Chile, Santiago, Chile
b Department of Industrial Engineering, Universidad del Bío-Bío, Concepción, Chile
c Department of Management and Information Systems, Universidad de Chile, Santiago, Chile
d Department of Engineering Sciences, Universidad Andres Bello, Santiago, Chile

Article history:
Received 30 July 2014
Received in revised form 28 September 2015
Accepted 10 May 2016
Available online 11 May 2016

Keywords:
Operating room scheduling
Emergency surgeries
Hybrid flow shop
Constraint programming
Genetic algorithm

ABSTRACT

Surgery rooms are among the most expensive resources in hospitals and clinics. Their scheduling is difficult because, in addition to the surgical room itself, each surgery requires a particular combination of human resources, as well as different pieces of equipment and materials. Furthermore, after each surgery, a post-anesthesia bed is required for the patient to recover. Finally, in addition to planned surgeries, the scheduling must be made in such a way as to accommodate the emergency surgeries that may arrive during each day, which must be attended within a limited time. We address the surgery scheduling problem considering simultaneously, for the first time, the operating rooms, the post anesthesia recovery, the resources required by the surgery and the possible arrival of emergency surgeries. We propose an integer linear programming model that allows finding optimal solutions for small size instances, we transform it to use constraint programming, and develop a metaheuristic based on a genetic algorithm and a constructive heuristic, that solves larger size instances. Finally, we present numerical experiments.

1. Introduction

The relevance of the problem of planning and scheduling of operating rooms (ORs) comes from the fact that ORs are among the most expensive and fundamental resources of a hospital (Cardoen, Demeulemeester, & Beliën, 2010; Guerriero & Guido, 2011). Among other consequences, poor scheduling may generate idle time intervals, an excess of overtime, and delays or cancellations of surgeries, that are reflected in excess cost and loss of revenue for the hospital.

The process of scheduling of surgeries in ORs has two main stages: the daily assignment of patients and the appropriate sequencing of the surgeries (Cardoen, Demeulemeester, & Beliën, 2009a; Guerriero & Guido, 2011). The first stage consists of determining the set of patients that should be operated on a particular day over a given planning horizon (Guinet & Chaabane, 2003). Patients that are candidates for surgery are generally selected from patient waiting lists that are generated by the hospital (Persson & Persson, 2009). This selection is based on resource availability, surgery priority, and the patient’s waiting time in the waiting list. At this point, a surgeon is assigned to the surgery. In the second stage, the daily programming of each operating room is prepared, which is known as the Surgical Table Programming or Operating Table (OT). This activity determines the sequencing of the surgeries in the ORs and is scheduled 24–48 h before the event (May, Spangler, Strum, & Vargas, 2011). In this paper, we concentrate on the second step of the process of surgery scheduling, that is the sequencing of surgeries in the ORs.

The OT programmer should build the sequence taking into account the ORs availability as well as the physical and human resources needed to perform each surgery. At the same time, the programmer must consider the availability of a post-anesthesia recovery bed for the patient, thus making sure that the surgical room is not blocked, which would impact the following surgery. Not considering any of these physical or human resources may lead to the cancellation of scheduled surgeries.

Normally, the OT programmer is a physician or a nurse, familiar with the process of each of the scheduled surgeries. Nevertheless, even though the programmer is familiar with the elements, time and care needed to perform the surgery, achieving a good schedule is a very difficult process when a large number of factors need to be simultaneously addressed. In addition, every day, a number of

* Corresponding author.
E-mail address: marianov@ing.puc.cl (V. Marianov).

http://dx.doi.org/10.1016/j.cie.2016.05.016
0360-8352/© 2016 Elsevier Ltd. All rights reserved.
emergency surgeries arrive to the hospital which must be handled as soon as possible. Thus, even when surgeons are very skilled in performing surgery, large amounts of idle time occur due mainly to poor scheduling.

Our main contribution is the construction of the OT simultaneously considering for the first time the operating rooms, resources, post-anesthesia recovery beds, and emergency surgeries. We propose different tools to solve the problem and assess the results using computational tests.

The remainder of the article is organized as follows: Section 2 reviews the literature. Section 3 describes the problem. Section 4 presents the exact solution approaches and the tools proposed to solve of the problem. Section 5 describes the heuristic. Section 6 presents and describes the computer experiments. Finally, in Section 7 we offer conclusions and outline directions for future work.

2. Literature review

The scheduling of ORs is a broad topic. A detailed review can be found in Cardoen et al. (2010) and Guerriero and Guido (2011). Here, we focus on the literature that is directly related. Among the relevant papers are Jebali, Hadj Alouane, and Latot (2006) and Dekhici and Belkadi (2010), in which the assignment and sequencing of surgeries are treated as a hybrid flow shop (HFS), without considering any of the required additional resources in the scheduling process. Roland, Di Martinelly, and Riane (2006) and Roland, Di Martinelly, Riane, and Pochet (2010) emphasize the importance of considering additional resources, specially human, in the surgery scheduling process but the post-anesthesia recovery stage is not taken into account.

In Pham and Klinkert (2008), each surgical intervention is described as a predetermined sequence of activities with a maximum allowed waiting time between two consecutive activities, with a set of resources being assigned to each of the activities. The authors develop a mixed linear integer programming ( MILP) model. Vijayakumar, Parikh, Scott, Barnes, and Gallimore (2013) consider the resources in the problem and use a dual bin-packing problem model. Given the number of days and resources available for the schedule, they maximize the number of surgeries to schedule. They model the problem as a MILP and develop a heuristic to solve the problem without considering the second stage (recovery). The surgical working affinity group (group of surgeons, nurses and personnel that usually work together) is included as a limitation in Meskens, Duvivier, and Hanse (2013), and constraint programming (CP) is used to solve the problem. Lee and Yih (2014) determine the start times of the surgeries in the operating room based on the availability of beds in the unit Post-Anesthesia Care Unit (PACU) and the uncertainty of the duration of tasks. The problem is approached as a flexible job shop with fuzzy times. To solve it, they develop a genetic algorithm that determines the order of the surgeries and a heuristic to determine the start times.

Emergency surgeries cannot be programmed in advance, since they may arrive at any moment of the day. However, they must be attended promptly. To address this type of uncertainty, Erdem, Qu, and Shi (2012) formulate a MILP model and a genetic algorithm that allows rescheduling elective surgeries. The proposed model minimizes the cost of postponing surgeries, the overtime, and the cost of rejecting an emergency patient by deriving her to a different hospital. Wullink et al. (2007) use simulation to evaluate reservation of time for emergency surgeries. They evaluate two traditional methodologies: (a) concentrating all reserve capacity into an ORs dedicated exclusively to emergency surgery, and (b) uniformly reserving capacity in all ORs dedicated to elective surgery. The performance measures are waiting time, overtime, and OR utilization. They conclude that the second approach is the most efficient. van Essen, Hans, Hurink, and Oversberg (2012) study to minimize the emergency surgery waiting time, for which they introduce two new concepts: the “break-in-moment” (BIM) that corresponds to the exact moment in which an emergency surgery can be performed after the conclusion of an elective surgery, and the “break-in-interval” (BII), defined as the interval between two consecutive BIMs. What they look for is distributing BIMs uniformly throughout the day, for which they minimize the maximum BII. The authors assume that surgeries are programmed without dead times in between. In all these works, they use the fact, recognized in the literature, that there exists a so-called ‘golden hour’: a one-hour period within which care must be provided to trauma emergency patients (Fleet & Poitr, 2011; Newgard et al., 2010).

We synthesize the main literature in the following Table 1, summarizing the related works. The first column lists the authors, while the second to seventh columns show the features of the studied problem. As Table 1 shows, most of the papers partially address the features involved in the problem. Our model covers most of the features: programming for minimum makespan and consideration of the waiting time of emergency surgeries, sequencing and individual assignment of all the medical staff as well as all the resources, and consideration of the two important stages: the OR and the recovery bed.

3. Problem description

A patient’s surgery process consists of two main stages. The first stage makes use of the surgery room and it is divided into the OR preparation for surgery, the surgery itself (surgical act), and the cleaning of the OR. The second stage is the post-anesthesia recovery recovery of the patient that takes place in a recovery room.

In general, ORs have different sizes and characteristics, thus a surgery can only be assigned to an operating room the meets the conditions required to perform the operation. Priorities among surgeries may also occur. For example, it may be necessary to move towards the end of the day those surgeries that may contaminate the room to a larger degree.

Each surgery requires a previously assigned physician or surgeon. The surgeon is only present during the surgery act and requires a certain amount of time between interventions (mainly for cleaning, change of clothes, and rest) that depends on the duration of the previously performed surgery.

During a surgery additional resources are required, often scarce. These resources may be nurses, anesthesiologists and other professionals, or physical resources (instruments, imaging equipment, etc.). One or more units of a particular type of resource (human or physical) may be available. Each resource unit assigned to a surgery must be available throughout the preparation and performance of the surgery and has a fixed preparation time that depends on the resource type.

Once the surgery is finished, the patient must be transported to a bed of PACU. Some patients, after surgery, require special care and must be transported to an Intensive Care Unit (ICU). Thus, post-anesthesia recovery may require a special bed.

In addition, demand for emergency surgeries may occur at any time and must be attended within a maximum waiting time.

The objective of the problem is to minimize the closing time of the last operating room in use (makespan).

4. Proposed exact approaches

The problem of scheduling operating rooms, resources and post anesthesia care units can be seen as a hybrid flow shop (HFS) scheduling problem. This is an NP-Hard combinatorial problem
Gupta, Xie, and Perdomo (2010) also consider a two-stage environment, with one stage consisting of a machine, and adds to the final result. A simple flow shop is an environment in which products, tasks, or jobs, require processing involving more than one stage. Each stage consists of a machine, and adds to the final result. A hybrid flow shop has more than one (parallel) machine in at least one of the stages, and all machines in a stage can perform either the same task, or have different capabilities. Each machine can handle only one product, task or job at a time, and each task is processed by only one machine in each stage.

Our first approach is addressing the problem as a two-stage HFS. Each job (surgery) is assigned to exactly one machine, belonging to the set of parallel machines in stage 1 (ORs), and exactly to one machine belonging to the set of parallel machines at stage 2 (beds of PACU). In both stages, there is a subset of machines that can perform the job. Each job requires additional resources to be processed. There are setup times that are sequence-dependent, depending on the specific setup that needs to be performed at the beginning of the day, between surgeries or at the end of the day. The double arrow indicates the maximum waiting time for an emergency surgery (one hour). Note that this limit is not satisfied after surgery c starts. If an emergency surgery appears, once c has begun, it must wait until total completion of c. The completion of the cleaning after surgery b in OR 1 determines the makespan.

To solve the problem, we formulate a MILP. We transform it into a Constraint Programming model, as this methodology has shown to produce good quality feasible solutions with low computational cost.

4.1. Mathematical milp model

The notation for the model is as follows.

<table>
<thead>
<tr>
<th>Indexes</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>i, j</td>
<td>surgeries</td>
<td>k</td>
<td>ORs and beds of PACU</td>
</tr>
<tr>
<td>h</td>
<td>stages</td>
<td>a</td>
<td>surgeons</td>
</tr>
<tr>
<td>r</td>
<td>resources</td>
<td>n</td>
<td>total number of surgeries scheduled</td>
</tr>
</tbody>
</table>

The variable $s_{ij}$ represents the start time of surgery $i$ in stage $j$. Therefore, we can minimize the makespan by using the following objective function:

$$\min \sum_{i=1}^{n} C_i$$

where $C_i$ is the completion time for surgery $i$.

subject to the following constraints:

- **Resource Constraints**
  $$\sum_{i=1}^{n} s_{ij} + \sum_{k=1}^{k} r_{ik} \leq C_i$$

- **Delay Constraints**
  $$C_i - C_{i-1} \geq 0$$

- **Waiting Constraints**
  $$C_i - C_{i-1} - t_{ij} \leq 0$$

- **Precedence Constraints**
  $$C_i = 0$$

- **Start Time Constraints**
  $$s_{ij} = 0$$

Finally, there are urgent jobs (emergency surgeries) that can arrive, which must be attended within a standard amount of time. To address the emergency surgery issue, based on the results of Wullink et al. (2007), we do not consider assigning special ORs for such surgeries. Rather, we use the approach proposed in van Essen, Hans, et al. (2012), consisting in minimizing the maximum waiting time for an emergency surgery, and we transform it into a problem constraint, whereas we limit the maximum waiting time to a pre-determined value (60 min based on the ‘golden rule’).

Fig. 1 presents an example in which 3 surgeries (a, b, and c) are scheduled in 2 ORs and 2 beds of PACU. Note that the start of surgery c is delayed at stage 1, to have bed availability in stage 2 without a waiting period. The surgeon assigned to surgery c is only present during the surgical act, while the resource used (a nurse, for example) is present since the preparation of the surgery starts until it finishes. If emergency surgeries are required, they can be performed at the beginning of the day, between surgeries or at the end of the day. The double arrow indicates the maximum waiting time for an emergency surgery (one hour). Note that this limit is not satisfied after surgery c starts. If an emergency surgery appears, once c has begun, it must wait until total completion of c. The completion of the cleaning after surgery b in OR 1 determines the makespan.

To solve the problem, we formulate a MILP. We transform it into a Constraint Programming model, as this methodology has shown to produce good quality feasible solutions with low computational cost.
duration time (h = 1 is surgical act, h = 2 is post-anesthesia recovery) of surgery j

setup time of surgery j

clean time of surgery j

setup time of surgeon a, after to performing the surgery j

maximum waiting time for emergency surgery

numbers of ORs/beds (h = 1 is ORs, h = 2 is beds)

numbers of resource r

large integer number

set of surgeries

set of ORs/beds (h = 1 is OR, h = 2 is bed) where can be assigned to the surgery j

set of surgeries that require the resource

set of ORs/beds (h = 1 is OR, h = 2 is bed) assigned to the OR or bed

set of resources required for surgery j

set of surgeries that require the resource r

set of surgeries that cannot precede surgery j

The following are the variables of the model. These variables specify the allocation of surgeries to operating rooms in stage 1 and post-anesthesia beds in stage 2. Also, they allow allocating resources and scheduling in such a way that the maximum waiting time for emergency surgeries, $E_{\text{max}}$ is not exceeded.

$x_{jh}^k = 1$ if surgery $j$, in stage $h$ ($h = 1$ is OR, $h = 2$ is bed), is assigned to the OR or bed $k$, 0 otherwise.

$y_{ji}^h = 1$ if the end of surgery $i$ precedes the end of surgery $j$ in stage $h$, 0 otherwise.

$w_{ij}^r = 1$ if the start of surgery $i$ precedes the end of surgery $j$ and both surgeries require of resource $r$, 0 otherwise.

$w_{ij}^{pr} = 1$ if surgery $i$ and surgery $j$ are performed in parallel at an instant of time and both surgeries require of resource $r$, 0 otherwise.

$z_{ij}^h = 1$ if the gap between the start of surgery $i$ and the end of surgery $j$ is greater than $E_{\text{max}}$, 0 otherwise.
\[ C_i^j + L_i \leq C_j^i + L_j + G_i(1 - z_{ij}^0), \quad \forall j \in J; i \in J i \neq j \] (14)

\[ z_{ij}^0 + z_{ji}^0 = 1, \quad \forall j \in J; i \in J i \neq j \] (15)

\[ E_{\text{max}} \geq (C_i^j + L_j) - (C_i^j - P_i^j - S_i) - z_{ij}^0 G_i, \quad \forall j \in J; i \in J i \neq j \] (16)

\[ z_{ij}^0 + z_{ji}^0 \leq z_{ij}^0 + z_{ji}^0, \quad \forall i \in J : P_i^j + S_i + L_j > E_{\text{max}} \] (17)

\[ \sum_{j=1}^{m} z_{ij}^0 \leq m^1 - 2, \quad \forall i \in J : P_i^1 + S_i + L_i > E_{\text{max}} \] (18)

\[ C_{\text{max}} \geq C_i^j + L_j, \quad \forall j \in J \] (19)

\[ x_{rik}^0 \in \{0, 1\}, \quad \forall h = 1, 2; j \in J; k \in K_h^i \] (20)

\[ y_{kij}^0 \in \{0, 1\}, \quad \forall h = 1, 2; j \in J; i \in J i \neq j \] (21)

\[ w_{rik} = w_{rik}^2 \in \{0, 1\}, \quad \forall r \in U; j \in J; i \in J i \neq j \] (22)

\[ z_{ij}^1, z_{ijk}^2, z_{ik}^1 \in \{0, 1\}, \quad \forall j \in J; i \in J i \neq j \] (23)

\[ C_i^0 \geq 0, \quad \forall h = 1, 2; i \in J \] (24)

\[ C_{\text{max}} \geq 0. \] (25)

The objective function (1) minimizes the closing time of the last OR in use. Constraint (2) requires all surgeries being assigned to only one OR and one bed of PACU. Constraint (3) states that the surgery \( i \) precedes surgery \( j \) in stage \( h \), if the completion time of \( i \) is lesser than or equal to the completion time of surgery \( j \) in stage \( h \). Constraint (4) requires a single predecessor for each combination of two surgeries. Constraint (5) sets the completion time of any surgery to a value greater or equal than the sum of the setup time of surgeries that are assigned to a same OR. Constraint (6) determines the precedence between surgeries that are assigned to a same bed of PACU. Constraint (8) forces no waiting times to move from stage 1 to stage 2, whereas (9) enforces priorities between operations. Constraint (10) determines the order in which each surgeon performs her/his assigned surgeries. Constraint (11) forces the start of surgery that uses a determined resource to never start before the completion of another surgery using the same resource. Constraint (12) avoids surgeries using the same resource to be performed at the same time or in parallel, while (13) states that the maximum number of surgeries programmed in parallel is given by the number of available resources. Constraint (14) determines the precedence between the completions of the cleaning activities after the surgeries. For each combination of surgeries, constraint (15) states that there is only one predecessor. Constraint (16) requires the difference between the end of surgery \( j \) and the start of surgery \( i \) to be lesser than or equal to \( E_{\text{max}} \). Constraint (17) determines whether the time of the surgeries scheduled in parallel is greater than \( E_{\text{max}} \). Constraint (18) ensures that \( E_{\text{max}} \) is met. Constraint (19) indicates that \( C_{\text{max}} \) is greater or equal to the closing time of any ORs. Finally, Constraints (20)-(25) define the domain of the variables.

### 4.2. Constraint programming model

The CP model is formulated using the same indexes, parameters and sets described for the MILP. Additionally, it uses the following variables:

\[ x_{ijk}^0 = \text{OR(\text{bed}) where is programmed the surgery } i \text{ in stage } j, \text{ where } h = 1 \text{ is OR, } h = 2 \text{ is bed}. \]

\[ C_{\text{startj}} = \text{start time of surgery } j \text{ in the stage } h. \]

\[ C_{\text{endj}} = \text{end time of surgery } j \text{ in the stage } h. \]

The CP uses the same objective function (1) as the MILP, and the following are the sets of constraints:

\[ C_{\text{startj}} + P_i^j + S_i = C_{\text{endj}}, \quad \forall j \in J \] (26)

\[ C_{\text{startj}} + P_i^j = C_j^2, \quad \forall j \in J \] (27)

\[ C_{\text{startj}} = C_{\text{endj}}, \quad \forall j \in J \] (28)

if \( x_{ijk}^0 = x_{ijk}^1 \) then \( C_i^1 + L_i \leq C_{\text{startj}} \lor C_{\text{endj}} + S_i < C_{\text{startj}} \lor \) \( \forall j \in J; i \in J i \neq j \) (29)

if \( x_{ijk}^1 = x_{ijk}^0 \) then \( C_{\text{endj}} \leq C_{\text{startj}} \lor C_{\text{endj}} < C_{\text{startj}} \lor C_{\text{endj}} \), \( \forall j \in J; i \in J i \neq j \) (30)

if \( x_{ijk}^2 = x_{ijk}^0 \) then \( C_{\text{endj}} + L_i \leq C_{\text{startj}} \lor \forall j \in B; i \in B_j \) (31)

\[ C_{\text{endj}} + S_i^g \leq C_{\text{startj}} + S_i \lor C_{\text{endj}} + S_i^g \leq C_{\text{startj}} + S_j, \quad \forall a \in M; j \in M^a; i \in M_a i \neq j \] (32)

\[ u_r - 1 \geq \sum_{\forall r, j} \left( \left( C_{\text{startj}} < C_{\text{endj}} + S_i \lor C_{\text{startj}} < C_{\text{endj}} + S_j \lor C_{\text{endj}} \right) \right. \]

\[ \left. \leq C_{\text{endj}} \text{ then } 1 \right) \lor \forall r \in U; j \in J_i \] (33)

\[ m^1 - 2 \geq \sum_{\forall r, j} \left( \left( C_{\text{endj}} + L_i - C_{\text{startj}} > E_{\text{max}} \lor C_{\text{endj}} \leq C_{\text{startj}} \text{ then } 1 \right) \right. \]

\[ \left. \lor j \in J : P_i^j + T_j + L_j > E_{\text{max}} \right) \] (34)

\[ C_{\text{endj}} + L_j \leq C_{\text{max}}, \quad \forall j \in J \] (35)

\[ x_{ijk}^0 \in K_h^i, \quad \forall j \in J; h = 1, 2 \] (36)

\[ C_{\text{startj}} + C_{\text{endj}} \geq 0, \quad \forall h = 1, 2; j \in J \] (37)

The set of constraints in (26) states that the surgery end time is its start time plus the duration of the preparation and the surgical act duration. Constraint (27) indicates that the completion time of surgery in stage 2, is the start of the post-anesthesia recovery plus the recovery time. Constraint (28) is the no wait constraint between stages one and two, for any surgery. Constraint (29) indicates that, if two operations are scheduled in the same OR, one must precede the other. Constraint (30) is similar to (29), for the recovery phase. Constraint (31) sets priorities between surgeries. Constraint (32) indicates that a surgeon can only perform one surgery at a time. Constraint (33) indicates that the number of units of a resource used simultaneously depends on the available units for that resource. Constraint (34) restricts the maximum waiting time for emergency surgeries. Constraint (35) forces \( C_{\text{max}} \) to be equal to
the closing time of the latest ORs. Finally, Constraints (36), (37) declare the domain of the decision variables.

5. Metaheuristic

Due to the complexity of the problem (it is NP Hard), a metaheuristic is proposed, based on a genetic algorithm (GA) and a constructive heuristic (CH), which finds feasible solutions to large-scale instances within a reasonable execution time.

The main idea is as follows: the GA generates a population consisting of vectors representing possible sequences in which the surgeries could be assigned to ORs. For each sequence, the CH computes the corresponding values of the starting and ending times of all the surgeries (vector elements), assigns resources and post-anesthesia beds to them according to availability, checks for feasibility, and computes the sequence's objective value. Each GA-generated sequence, completed with the values computed by the CH, is a feasible solution to the problem. In order to improve the quality of the population, the GA replaces old solutions with new ones, if these have a better objective value. Once the stopping condition is met, the best solution is selected among all the solutions in the population.

5.1. Genetic algorithm

The GA generates an initial population by randomly performing permutations on a vector containing all the surgeries. Each permutation or GA generated solution represents a sequence in which surgeries can be assigned to ORs.

In each iteration, the GA randomly chooses two solutions from the current population and crosses or combines them to generate two new solutions. One of these new solutions is selected as a candidate to replace an old solution. With a low probability, the GA applies a mutation operator to this candidate solution, consisting of trying all exchange options between any two of its elements (surgeries) and evaluating the objective each time. The best resulting solution is stored and becomes the new candidate solution. The candidate solution replaces the worst solution in the current population if its objective value is better, maintaining a fixed size population.

We use two stopping conditions: a number of consecutive iterations after which there has been no improvement in the value of the objective function and a maximum of total iterations.

The crossing between two solutions in the reproductive stage of the GA is the linear order crossover described in Falkenauer and Bouffoux (1991). Let vectors \( \text{vect}_1 \) and \( \text{vect}_2 \) be the solutions selected for the crossover and \( \text{vect}_3 \) and \( \text{vect}_4 \) the resulting solutions after the crossover. Two cut points within vectors \( \text{vect}_1 \) and \( \text{vect}_2 \) are selected randomly, \( \text{cut}1 \) and \( \text{cut}2 \). All elements of \( \text{vect}_1 \) between these cuts are preserved in \( \text{vect}_3 \), in the same positions. The remaining elements of \( \text{vect}_3 \) are filled using elements of \( \text{vect}_2 \), as follows: pick the element of \( \text{vect}_2 \) to the right of \( \text{cut}2 \) and replace the same element of \( \text{vect}_3 \). If the surgery represented by that element has not been already included in \( \text{vect}_3 \), iterate taking each time the next element to the right in both vectors, and when the end of the vector is reached, do the same starting from the first element until \( \text{cut}1 \) is reached. In the example of Fig. 2, the elements between \( \text{cut}1 \) and \( \text{cut}2 \), third and fourth elements of \( \text{vect}_1 \), are preserved in \( \text{vect}_3 \). Then, starting from \( \text{cut}2 \), the value at the right of \( \text{cut}2 \) should replace the same position of \( \text{vect}_4 \), but since the new value already belongs to \( \text{vect}_3 \), the procedure moves to the next element to the right in the origin vector \( \text{vect}_2 \). In this case, there are no more elements to the right in \( \text{vect}_2 \), and the procedure jumps to the first element of \( \text{vect}_2 \), which is \( d \). Since surgery \( a \) is already contained in \( \text{vect}_3 \), the next element in \( \text{vect}_2 \), \( b \), is used to fill the position in \( \text{vect}_4 \). The following elements of \( \text{vect}_4 \) are filled with the next elements of \( \text{vect}_2 \), \( c \) and \( d \).

5.2. Constructive heuristic

The CH performs all the calculations required to compute the value of the objective function for each sequence of surgeries generated by the GA. The CH consists of the following 7 steps. The first 6 steps are applied to each surgery or element of the sequence, while the seventh step is applied once the 6 first steps have been applied to all surgeries in the sequence:

Step 1: Starting Time of Surgery. Assign the surgery to the least used OR. Set the starting time of the surgery.

Step 2: Availability of Surgeon. Delay (if required) the starting time of the surgery just assigned, to meet the soonest time at which the surgeon is available.

Step 3: Availability of Resources. Delay (if required) the starting time of the surgery to meet the soonest time at which all the resources are available.

Step 4: Maximum waiting time. Delay (if required) the starting time of the surgery to meet \( E_{\text{max}} \).

Step 5: Availability of Bed in the PACU. Delay (if required) the starting time of the surgery to meet the sooner time at which a post-anesthesia bed is available.

Step 6: Time Allocation. Compute the occupancy time for all resources (OR, bed, nurse, etc.).

Step 7: Compute \( C_{\text{max}} \). If there are violated priorities between surgeries, repair the scheduling and recalculate the times. Otherwise, compute \( C_{\text{max}} \).

5.3. Complete metaheuristic

Fig. 3 presents the pseudo-code of the complete metaheuristic, including GA and CH.

6. Computational experiments

The master group of surgeries shown in Table 2 was generated randomly. The master group includes information on the clinical specialty to which each surgery belongs, the resources required at the time of the operation and the assigned surgeon.

The duration of the surgeries in master Table 2, shown in Table 3, was generated randomly using a log-normal distribution, along with the information presented in Table 3, taken from Marcon, Kharraja, Smolski, Luquet, and Viale (2003). To generate setup times and cleaning times, the methodology of Jebali et al. (2006) was used. If the duration of surgery was less than 90 min, the preparation time of the OR was set to 10 min, otherwise to 20 min. The preparation and anesthesia time of the patient was 10 min if the surgical act lasted between 60 and 120 min, and 30 min otherwise. If the duration of surgical act was less than 20 min, the OR cleaning lasted 15 min, 30 min otherwise. The recovery time post-anesthesia is generated using a log-normal distribution with a mean equal to the duration of surgery minus 10 min, and a standard deviation of 15 min (Jebali et al., 2006).

Furthermore, if the time of surgical act was less than 60 min, a time of preparation of the surgeon was set to 15 min. If the time of surgical act was between 60 and 120 min, the preparation time of the surgeon was 30 min, and 45 min otherwise. Preparation time of resources was generated randomly between 15 and 90 min.

All times were rounded to multiples of 5. For example, if the random time value obtained is 33.8 min, it was rounded to 35 min. This is consistent with what is done in practice. The
The number of ORs used was generated as the sum of the time of preparation, cleaning and surgical act divided by 600 min (10 h), i.e.
\[ m_1 = \frac{1}{600} \sum_{j} (L_j + T_j + P_j) \]. The number of post-anesthesia beds was generated randomly, varying in the range \( m_1 / C_0 \leq m_2 \leq 2m_1 \). The number of resource units was generated randomly varying in the range \( 1 \leq u_r \leq Jr \).

30 instances were generated using the surgeries in the master set. These instances consider 15, 16, 17, 19, 21, 24, 27, 30, 35 and 40 surgeries, and for each number of surgeries, three instances were developed. For example, for instances 1, 2 and 3, 15 surgeries were selected randomly from Table 2, and so on. For instance 28, 29 and 30, all 40 surgeries were used.

Both MILP and CP models, including the heuristic, were programmed in C++, using Win32 Console Application of Visual Studio 2010. MILP and CP were solved with the CPLEX 12.6 package. All the algorithms were programmed with a run limit. The MILP was run with a time limit of 1 h (3600 s) and a limit of 8 h (28,800 s) for each instance. The CP was run with a limit of 1 h and 1,000,000 failures allowed before terminating a search. The heuristic was programmed to stop after 1500 iterations without changes.

Table 2
Master set of surgeries to be scheduled.

<table>
<thead>
<tr>
<th>Type of surgery</th>
<th>Id.</th>
<th>Surgery</th>
<th>Resources</th>
<th>Surgeon</th>
<th>Type of surgery</th>
<th>Id.</th>
<th>Surgery</th>
<th>Resources</th>
<th>Surgeon</th>
</tr>
</thead>
<tbody>
<tr>
<td>General</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>Ambulatory</td>
<td>21</td>
<td>1</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td></td>
<td>22</td>
<td>1</td>
<td>8</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>4</td>
<td></td>
<td>23</td>
<td>1</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>Orthopedics</td>
<td>24</td>
<td>1</td>
<td>8</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>4</td>
<td>6</td>
<td>6</td>
<td></td>
<td>25</td>
<td>1</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>3 5</td>
<td>6 7</td>
<td></td>
<td></td>
<td>26</td>
<td>1</td>
<td>8</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>1 8</td>
<td></td>
<td></td>
<td></td>
<td>27</td>
<td>1</td>
<td>8</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>1</td>
<td>8</td>
<td></td>
<td></td>
<td>28</td>
<td>1</td>
<td>8</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>1</td>
<td></td>
<td></td>
<td>Endoscopy and radiology</td>
<td>10</td>
<td>7 8</td>
<td>9</td>
<td>Otorhino-laryngology</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td>4</td>
<td>5</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>31</td>
<td>16</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
<td>32</td>
<td>16</td>
<td>17 18</td>
<td>23</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>8 9 10</td>
<td></td>
<td></td>
<td></td>
<td>33</td>
<td>17</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
<td>34</td>
<td>17</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>9</td>
<td></td>
<td></td>
<td>Ambulatory</td>
<td>35</td>
<td>16</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>9 10</td>
<td></td>
<td></td>
<td></td>
<td>36</td>
<td>4 18</td>
<td>19</td>
<td>25</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>2 11</td>
<td></td>
<td></td>
<td></td>
<td>37</td>
<td>4 18</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>14</td>
<td></td>
<td></td>
<td></td>
<td>38</td>
<td>4 18</td>
<td>25</td>
<td></td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>11 7</td>
<td></td>
<td></td>
<td></td>
<td>39</td>
<td>4 18</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td></td>
<td>20</td>
<td>12 13</td>
<td>16</td>
<td></td>
<td></td>
<td>40</td>
<td>4 18</td>
<td>27</td>
<td></td>
</tr>
</tbody>
</table>

Fig. 2. Example GA.

Fig. 3. Heuristic Pseudo-code.
in the population or a maximum of 10,000 total iterations. The computational tests were performed on a computer with operating system Windows 7 Professional 64-bit, a processor Intel (R) Core (TM) i7-2600 3.40 GHz CPU @ 3.40 GHz and 16 GB RAM. All tests were performed using only one processor core.

Table 4 shows some of the results of the computational tests using the MILP, CP and heuristic. The first column of Table 4 shows the number of the instance while the second shows the total number of scheduled surgeries considered in that instance. The objective value, run time and gap obtained with the MILP (both executed for 1 h), MILP gets better solutions than CP in 6 instances, while CP is better in 3 instances. The largest difference occurs in instance 8, when the CP overcomes the MILP.

The heuristic obtained very good solutions in a very short run time. It found the optimal solution in 18 of the 19 solutions shown to be optimal by the MILP. The heuristic even achieved better solutions than MILP with a time limit of 8 h in instances 13 and 21.

<table>
<thead>
<tr>
<th>Instance</th>
<th>Number of surgeries</th>
<th>MILP (3600 s)</th>
<th>MILP (28,800 s)</th>
<th>CP (Fail Limit)</th>
<th>CP (3600 s)</th>
<th>Heuristic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Obj. (min)</td>
<td>Time (s)</td>
<td>Gap (%)</td>
<td>Gap (%)</td>
<td>Obj. (min)</td>
<td>Time (s)</td>
</tr>
<tr>
<td>1</td>
<td>15</td>
<td>570</td>
<td>15.3</td>
<td>0.0</td>
<td>414.6</td>
<td>Feas</td>
</tr>
<tr>
<td>2</td>
<td>15</td>
<td>505</td>
<td>Limit 4.0</td>
<td>Limit</td>
<td>510</td>
<td>454.9</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>570</td>
<td>Limit 26.3</td>
<td>Limit 23.5</td>
<td>590</td>
<td>393.5</td>
</tr>
<tr>
<td>4</td>
<td>15</td>
<td>545</td>
<td>Limit 37.1</td>
<td>Limit 30.7</td>
<td>560</td>
<td>599.9</td>
</tr>
<tr>
<td>5</td>
<td>16</td>
<td>515</td>
<td>Limit 17.5</td>
<td>Limit</td>
<td>525</td>
<td>631.0</td>
</tr>
<tr>
<td>6</td>
<td>16</td>
<td>525</td>
<td>Limit 17.1</td>
<td>Limit</td>
<td>530</td>
<td>444.0</td>
</tr>
<tr>
<td>7</td>
<td>17</td>
<td>560</td>
<td>Limit 1.0</td>
<td>Limit</td>
<td>575</td>
<td>680.0</td>
</tr>
<tr>
<td>8</td>
<td>17</td>
<td>585</td>
<td>Limit 2.5</td>
<td>Limit 21.8</td>
<td>561</td>
<td>538.0</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td>625</td>
<td>9.6</td>
<td>0.0</td>
<td>4.1</td>
<td>Opt</td>
</tr>
<tr>
<td>10</td>
<td>19</td>
<td>685</td>
<td>14.4</td>
<td>0.0</td>
<td>1.6</td>
<td>Opt</td>
</tr>
<tr>
<td>11</td>
<td>19</td>
<td>670</td>
<td>6.2</td>
<td>0.0</td>
<td>2.5</td>
<td>Opt</td>
</tr>
<tr>
<td>12</td>
<td>19</td>
<td>565</td>
<td>Limit 13.4</td>
<td>Limit 575</td>
<td>406.3</td>
<td>14.9</td>
</tr>
<tr>
<td>13</td>
<td>21</td>
<td>605</td>
<td>Limit 16.5</td>
<td>Limit 615</td>
<td>945.0</td>
<td>17.9</td>
</tr>
<tr>
<td>14</td>
<td>21</td>
<td>665</td>
<td>Limit 22.6</td>
<td>Limit 20.2</td>
<td>655</td>
<td>1060.5</td>
</tr>
<tr>
<td>15</td>
<td>21</td>
<td>605</td>
<td>3.6</td>
<td>0.0</td>
<td>0.9</td>
<td>Opt</td>
</tr>
<tr>
<td>16</td>
<td>24</td>
<td>770</td>
<td>54.5</td>
<td>0.0</td>
<td>1.6</td>
<td>Opt</td>
</tr>
<tr>
<td>17</td>
<td>24</td>
<td>560</td>
<td>Limit 120.1</td>
<td>Limit</td>
<td>8.1</td>
<td>Opt 8.1</td>
</tr>
<tr>
<td>18</td>
<td>24</td>
<td>600</td>
<td>58.8</td>
<td>0.0</td>
<td>6.4</td>
<td>Opt</td>
</tr>
<tr>
<td>19</td>
<td>27</td>
<td>830</td>
<td>77.0</td>
<td>0.0</td>
<td>1790.6</td>
<td>Feas 1790.6</td>
</tr>
<tr>
<td>20</td>
<td>27</td>
<td>705</td>
<td>507.5</td>
<td>0.0</td>
<td>2566.4</td>
<td>Feas 2566.4</td>
</tr>
<tr>
<td>21</td>
<td>27</td>
<td>575</td>
<td>Limit 40.9</td>
<td>Limit 38.7</td>
<td>571</td>
<td>1704.7</td>
</tr>
<tr>
<td>22</td>
<td>27</td>
<td>705</td>
<td>143.2</td>
<td>0.0</td>
<td>2801.2</td>
<td>Feas 2801.2</td>
</tr>
<tr>
<td>23</td>
<td>30</td>
<td>1165</td>
<td>311.6</td>
<td>0.0</td>
<td>5562.3</td>
<td>Feas 5562.3</td>
</tr>
<tr>
<td>24</td>
<td>30</td>
<td>975</td>
<td>39.8</td>
<td>0.0</td>
<td>2078.4</td>
<td>Feas 2078.4</td>
</tr>
<tr>
<td>25</td>
<td>35</td>
<td>765</td>
<td>320.2</td>
<td>0.0</td>
<td>22.6</td>
<td>Opt 22.6</td>
</tr>
<tr>
<td>26</td>
<td>35</td>
<td>865</td>
<td>145.9</td>
<td>0.0</td>
<td>15.5</td>
<td>Opt 15.5</td>
</tr>
<tr>
<td>27</td>
<td>35</td>
<td>945</td>
<td>102.6</td>
<td>0.0</td>
<td>3450.5</td>
<td>Feas 3450.5</td>
</tr>
<tr>
<td>28</td>
<td>40</td>
<td>815</td>
<td>1204.4</td>
<td>0.0</td>
<td>5945.3</td>
<td>Feas 5945.3</td>
</tr>
<tr>
<td>29</td>
<td>40</td>
<td>780</td>
<td>472.9</td>
<td>0.0</td>
<td>21.5</td>
<td>Opt 21.5</td>
</tr>
<tr>
<td>30</td>
<td>40</td>
<td>720</td>
<td>1072.0</td>
<td>0.0</td>
<td>3378.4</td>
<td>Feas 3378.4</td>
</tr>
</tbody>
</table>

* Identical values to those obtained by the MILP with a time limit of 3600 s.
* Gap calculated using the MILP bounds with a time limit of 28,000 s.
For comparison purposes, the problem was solved in three relaxed versions: without consideration of the emergency surgeries, i.e., without caring for having intervals of at most one hour; without the second stage (recovery), and without resources. By not considering emergency surgeries, 47% of the solutions violate the \( E_{\text{max}} \) constraint. When it did not consider the recovery beds, 74% of the instances had a schedule conflict with respect to the availability of beds. By not considering the resources, 100% of the solutions had conflicts regarding resource demand. These relaxed solutions show the relevance of considering all these factors together.

As the programming in our method was obtained using average surgery times, we analyzed the robustness of the solution considering the uncertainty in the duration of the surgeries. We generated 2000 realizations of the duration of the surgeries, using a Monte Carlo method. The arrival times of up to 4 emergency surgeries in a day were assumed uniformly distributed between 0 and \( C_{\text{max}} \). The results were analyzed, keeping fixed the scheduling obtained by the MILP, CP model or heuristics. The analysis focused on the delay of each surgery with respect to its scheduled starting time, the waiting time that an emergency surgery can suffer in excess of \( E_{\text{max}} \), and the delay in the closing time of the last OR (\( C_{\text{max}} \)). Table 5 shows the mean and standard deviation of the results for five instances of different sizes.

As Table 5 shows, when considering uncertainty in surgery times the average closing time is longer than the value obtained deterministically in up to approximately one hour, with a standard deviation that is also close to one hour. As the number of scheduled surgeries increases, the delay in relation to the scheduled start time of each surgery tends to increase. The waiting time of emergency surgeries barely exceeds the maximum time \( E_{\text{max}} \). This is because the programming distributes short surgeries among the different ORs and in time, avoiding blocking simultaneously all ORs for too long times. Moreover, as the size of the instance grows, so does the number of ORs, increasing the probability of having an OR available on arrival of an emergency surgery.

### 7. Conclusions

Efficient scheduling of operating rooms is extremely difficult, and any improvement in the process could mean significant savings. Usually, this problem is solved sequentially, i.e., surgeries are assigned to ORs in a first stage. In a second stage, the equipment and other resources are assigned to the surgeries and, finally, the recovery beds are added to the programming. Solving this problem sequentially does not allow finding the bottlenecks in the programming until the last stage is solved and the recovery beds assigned. In this paper we develop several tools to improve the scheduling, that address the issue more comprehensively, while addressing all stages at once, considering not only the assignment of surgeries to operating rooms, but all the resources required for each surgery (human and material), and the recovery beds. Jointly addressing all these issues not only reduces bottlenecks, but allows finding a truly optimal solution to the problem. Furthermore, we include in the programming the emergency surgeries, through a deterministic solution that considers spacing the planned inter-surgery times by no more than one hour. This procedure assures prompt attention to any emergency surgery that might be required. The developed tools include a MILP model, a CP model and an ad-hoc heuristic, whose results are compared in run time and quality. Finally, we perform a probabilistic analysis of the problem, by generating a number of realizations and checking the performance of the procedure.

The CP model delivers good quality solutions in run times that are, in general, shorter than those of the MILP. However, the improvement is not significant. The metaheuristic, on the other hand, delivers solutions of very good quality in short run times, with averages of 7 s, on instances ranging from 15 to 40 surgeries requiring scheduling. The gap between MILP solutions and heuristic solutions never exceeded 3.6%, while in some instances, the solutions found by the heuristic were better than solutions obtained by using time-limited MILP in up to a 5.2%. With these short run times and good results, this heuristic could also be used for online surgery rescheduling, with small modifications.

A natural next step is explicitly including uncertainty in the duration of the activities and arrival of emergency surgeries.

### Acknowledgements

We thank the valuable comments made by a referee, which helped significantly improving the quality of the paper. Lüer-Villagra and Marianov gratefully acknowledge support by the Institute Complex Engineering Systems through grants ICM P-05-004-F and CONICYT FBO16.

### References


Table 5

<table>
<thead>
<tr>
<th>Instance</th>
<th>Number of surgeries</th>
<th>Surgery delay</th>
<th>Extra waiting time</th>
<th>Delay in ( C_{\text{max}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean</td>
<td>S.D.</td>
<td>Mean</td>
<td>S.D.</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td></td>
<td>13.1</td>
<td>3.03</td>
</tr>
<tr>
<td>9</td>
<td>17</td>
<td></td>
<td>19.4</td>
<td>41.7</td>
</tr>
<tr>
<td>17</td>
<td>24</td>
<td></td>
<td>23.4</td>
<td>46.5</td>
</tr>
<tr>
<td>22</td>
<td>30</td>
<td></td>
<td>23.1</td>
<td>52.7</td>
</tr>
<tr>
<td>30</td>
<td>40</td>
<td></td>
<td>28.9</td>
<td>53.4</td>
</tr>
</tbody>
</table>

determining start-times of surgical cases. European Journal of Operational

Research, 238(2), 620–629.


the number of beds in the postanesthesia care unit: A computer simulation flow


Portuguese hospital using a genetic heuristic. Operations Research for Health Care,
3(2), 59–72.

scheduling problem: Current research and future opportunities. Production and

Meskens, N., Duvivier, D., & Hanset, A. (2013). Multi-objective operating room
scheduling considering desiderata of the surgical team. Decision Support Systems,
59(2), 650–659.

... Nichol, G. (2010). Emergency medical services intervals and survival in
trauma: Assessment of the “golden hour” in a North American prospective

than anaesthetists to staff them: A stochastic simulation study. British Journal of
Anesthesia, 98(4), 462–469.

lagrangian relaxation. International conference on service systems and service
management (Vol. 2, pp. 1234–1239).


Pham, D. N., & Klinkert, A. (2008). Surgical case scheduling as a generalized job shop
scheduling problem. European Journal of Operational Research, 185(3),
1011–1025.

flow shop scheduling problems from a production system and a solutions

resource-constrained based solving approach. International conference on service
systems and service management (Vol. 1, pp. 443–448).

theatre under human resource constraints. Computers & Industrial Engineering,
58(2), 212–220.

Ruíz, R., & Vázquez-Rodríguez, J. A. (2010). The hybrid flow shop scheduling

waiting time for emergency surgery. Operations Research for Health Care, 1(2–3),
34–44.

van Essen, J. T., Hurink, J. L., Hartholt, W., & van den Akker, B. J. (2012). Decision
support system for the operating room rescheduling problem. Health Care

Vijayakumar, B., Parikh, P. J., Scott, R., Barnes, A., & Gallimore, J. (2013). A dual bin-
packing approach to scheduling surgical cases at a publicly-funded hospital.

Journal of Medical Systems, 31(6), 543–546.