CrossMark

Scheduling and rescheduling elective patients in operating rooms to minimise the percentage of tardy patients

Francisco Ballestín[1] · **Ángeles Pérez**[1] · **Sacramento Quintanilla**[1]

Abstract

This paper considers the problem of building a schedule of elective surgery patients for a unit of a hospital. Two phases are followed. In the first phase, a tentative schedule is calculated 2 weeks before the planning period looking to minimise the percentage of tardy patients. Several weight-based short-term objective functions are defined and compared to optimise this goal in the long term. In the second phase, a second and final schedule is made a few days before the planning period, taking into account the changes in the available information. We propose a general methodology to calculate this final schedule. Several strategies to manage the types and importance of changes from the tentative schedule are presented considering the idiosyncrasy of dealing with elective patients. The computational experiments have been carried out via simulation on instances randomly generated according to the data from a Spanish hospital.

Keywords OR in health services · Operating room · Scheduling

1 Introduction

Operating theatres or operating rooms are critical elements in hospitals. They are the units with the highest cost and the highest revenue; see e.g. Macario et al. (1995). Among the many decisions involving operating rooms, one of the most important is to build an effective and efficient surgery schedule. Several reviews have been written in the area; we refer to Cardoen et al. (2010), Hanset (2010), Guerriero and Guido (2011) and Meskens et al. (2013). As an example of other planning decisions in health care, consult the review of Hulshof et al. (2012).

In this paper, we study two optimisation problems that arise when selecting and scheduling patients waiting for surgery in a specific unit of a hospital. The first problem is to produce a tentative schedule for an established planning period, given the goals sought by the hospital and the available information at that moment. The main contribution of this part will be the study of which objective function can be used to obtain good results not only in the period considered, but also in a long-term period. The second problem considers the construction of a final schedule for the selected planning period based on the previously generated tentative schedule and taking into account the changes that have taken place in the days between the creation of the two schedules. Strategies to manage the changes from the tentative schedule are presented.

This work is part of a project in which we are analysing the benefits of incorporating optimisation techniques in several processes in one of the best and biggest public hospitals in Spain. The hospital serves an average of 4000 patients daily and has 1000 single beds, 45 operating rooms and 6300 employees. Our work is focused on the Urology unit, which is among the top performing units in the hospital regarding the treatment of patients on the waiting list. In the future, the aim is to study the application of these techniques to other units of the hospital.

The current objective of our work is to design a model that provides support for the selection and scheduling of patients waiting for surgery. This process is currently carried out by hand by the head doctor of the medical unit. We work with the so-called elective patients, whose surgery can be planned

✉ Francisco Ballestín
francisco.ballestin@uv.es

Ángeles Pérez
angeles.perez@uv.es

Sacramento Quintanilla
maria.quintanilla@uv.es

1 Dpto. de Matemáticas para la Economía y la Empresa,
Facultad de Economía, Universitat de València,
Avda Naranjos s/n, 46021 Valencia, Spain

well in advance. Currently, the Urology unit works with the so-called block (or blocked) scheduling strategy (see e.g. Fei et al. 2010). Surgeons or blocks of surgeons are assigned to a set of time blocks in which they can arrange their surgical cases. In our case, these time blocks correspond to specialties inside the urology unit, and only surgeries belonging to those specialties, apart from some surgeries called "general surgeries", can be performed in these slots. At this stage of our research, the assigned blocks are fixed. We are then interested in solving the sometimes called "surgery process scheduling problem". Let us describe the problem following the nomenclature of Agnetis et al. (2012).

For a given time period, decision-makers have to solve the following problems:

(i) to assign surgical disciplines to operating room sessions over time,
(ii) to assign elective surgeries to operating room sessions, and
(iii) to sequence these surgeries within the corresponding session.

Problem (i) is often referred to as the Master Surgical Schedule Problem (MSSP). Problem (ii) is denoted as the Surgical Case Assignment Problem (SCAP), whereas (iii) is the Elective Surgery Sequencing Problem (ESSP).

Our study works with a stable MSS, solving the SCAP and the ESSP. Due to the restrictions of our problem, the resolution of the ESSP is straightforward, so the paper focuses its attention on the SCAP.

To solve the SCAP, the head doctor of the unit calculates two schedules. The tentative schedule is created several days—in our case 2 weeks—before the planning period, with the information available at that moment. With this schedule, the patients to be operated on, and the doctors, are informed. A few days before the planning period, a second schedule is calculated to be used as the final schedule. The reason is that usually some changes in the data used for making the tentative scheduled occur. Some surgeries cannot be performed due to various reasons (unavailability of patients, doctors, necessary equipment, etc.), and new patients are available. This problem is called the rescheduling problem.

There are mainly two goals in the creation of the tentative schedule: the minimisation of the percentage of tardy patients and the maximisation of the utilisation rate of the ORs. Specifically, the hospital is satisfied if this rate is above a certain level: 80%. As we will see in the computational results in Sect. 4, it turns out that this level is clearly surpassed in the schedules generated by the proposed methods. Therefore, we will focus on the other goal.

Tardy patients are those who are scheduled after their due date. This due date is fixed by the local government based on the urgency of their illness. In machine scheduling, one talks about the fraction of tardy jobs or, sometimes, serviceability (Conway et al. 1967). The serviceability of a job indicates the probability that a job finishes on or before its due date. This is a goal that is rarely dealt with in the hospital literature, for several reasons. On the one hand, the goal cannot be directly modelled in a short-term model. On the other hand, this goal cannot be blindly followed. A hospital sticking to this goal would not operate a patient once he/she is tardy, something which is inacceptable. In this paper, we model this goal indirectly by means of short-term objective functions. Besides, these functions also try to obtain good results in the average tardiness, i.e. including tardy patients.

The review of Demeulemeester et al. (2013) groups papers according to different factors/fields, one of them being the performance criteria. The percentage of tardy patients is not mentioned as a major performance measure. Within the papers associated with other objective functions, there are a couple of papers worth mentioning. The objective function of Velasquez et al. (2008) includes a term that penalises the number of patients scheduled outside a given time window. The other term penalises the use of resources with additional capacity. The main difference with our model, apart from the second term in the objective function, is that in their model, conceived for a larger period, all patients have to be scheduled in that period. Therefore, the percentage of tardy patients can be directly included in the objective function. Persson and Persson (2009) solve a similar problem to ours. A law in Sweden states that patients must be operated on within 90 days or else receive paid surgery elsewhere. The paper suggests an approach that combines simulation and optimisation techniques for modelling surgery management decisions. We will follow a similar approach in this paper. The objective function also includes several elements. In the one concerning "tardy patients", the authors include a cost if a patient is not scheduled for surgery. This cost is the cost of scheduling the patient 8 weeks from the current planning period. Apart from these papers, Agnetis et al. (2012) use simulation to prove that introducing a very limited degree of variability in the assignment of OR sessions can pay off in terms of resource efficiency and due date performance. Due to the similarity of their goals with ours, we will use their objective function as a stepping stone.

Regarding the rescheduling problem, in the literature, similar problems have only been dealt with in a few papers and the objective is usually how to reschedule elective patients upon the arrival of emergency patients.

In Reilly et al. (1978), walk-in patients are rescheduled to a future appointment slot to cope with an overcrowded facility to improve the balance of resource utilisation over time. Li (2010) develop a simulation model to evaluate the efficiency of catheterisation laboratory operations in a major local health care facility. They vary the key parameters in the

model such as the length of the time block assigned to each case, the length of lunch buffers, as well as the option of rescheduling patients, and consider both operational costs and patient satisfaction metrics to illustrate the trade-offs between the two. Erdem et al. (2012) develop a mixed integer linear programming (MILP) model to solve this problem, by considering two types of clinical units, namely operating rooms and post-anaesthesia care units (PACUs). van Essen et al. (2012) and van Essen et al. (2013) develop a decision support system (DSS) which assists the OR manager in the decision when the planned OR schedule has to be adjusted due to surgery duration variability and/or arrivals of emergency surgeries. The DSS considers the preferences of all involved stakeholders and minimises the deviation from the preferences of the stakeholders. Li (2010) formulate the rescheduling problem as a variant of the bin-packing problem with interrelated items, which are the surgeries performed by the same surgeon. Bins are the staffed ORs and items are surgeries. The goal is to minimise the weighted sum of bins used, where the weight of a bin is proportional to its size. Addis et al. (2016) proposed a so-called rolling horizon approach for the patient selection and assignment. They calculate the schedule for several weeks with the given waiting list, leaving increasing free time in the schedule throughout the weeks for incoming patients. When unpredictable extensions of surgeries occur when applying the first week solution, some surgeries are cancelled and are rescheduled in the following weeks. Also, new arrivals are taken into account in the calculation of the final schedule. The midterm solution is rescheduled, limiting the number of variations from the previously computed plan.

Conforti et al. (2011) develop integer programming formulations to solve the scheduling of patients waiting for radiotherapy treatment. In the second of the two proposed model, it is possible to reschedule some patients. Rescheduling has here a double meaning: only the time slots can change with respect to the last planned week or it is possible to delay the day of the first weekly session. Luo et al. (2016) compare a non-rolling horizon model with a rolling horizon model in order to consider the variation in day-to-day demand, concluding that the average utilisation rate using the second model is significantly higher than using the first one. ValiSiar et al. (2017) schedule and reschedule patients, but also work with elective and semi-urgent patients. They use a rolling horizon scheduling–rescheduling framework. Their objective function takes into account three criteria: tardiness, overtimes and idle times.

None of these approaches serves our purpose. We need a first tentative complete schedule for the following two-week period with elective patients. Also, the final schedule is calculated taking into account the new information, but only works with elective patients. Similarly to Addis et al. (2016), changes are not welcome in general and therefore acceptance criteria and limits are imposed on them. However, we apply a method than do not impose a maximum specific number of changes

To sum up, the contribution of this paper is twofold. First, with regard to the choice of the objective function when building the tentative schedule, we are going to study which objective function in a short-term model leads to better results in the long-term goal used by the hospital (percentage of tardy patients). Comparisons among objective functions will be performed via simulation. A big computational effort has been put in the simulation. Tests have been performed over one year of plans with the corresponding updates of the waiting list with incoming and outcoming patients. We have followed the long-term evaluation method used among others by Agnetis et al. (2012) and called rolling horizon approach by Addis et al. (2016). It is further explained in Sect. 2.1.

The second contribution is the rescheduling of elective patients. Our methodology can be applied in many hospitals where a schedule of elective patients is calculated. It does not depend on the restrictions or the objective function. Furthermore, it does not depend on the number of operating rooms in the hospital or the number of surgeries performed by the unit. A distance between solutions is defined and used to identify the differences between two schedules and the importance of these differences to patients and doctors. We study the trade-off between the degree of the changes allowed in the tentative schedule and the benefits in the objective function obtained by them.

The rest of the paper is organised as follows: Sect. 2 introduces the model and the objective functions we propose to minimise the percentage of tardy patients. Section 3 discusses the rescheduling problem. Section 4 describes the instances we have used in the computational experiments, together with the simulation process and the computational results. Finally, the conclusions are given in the last section, i.e. Sect. 5.

2 Description of the problem. Short-term–long-term mathematical model

2.1 Description of the problem

There are three essential elements to consider in our optimisation problem: the operating room sessions assigned to the unit, the patients on the waiting list and the performance criteria. Let us describe each of them in detail. The medical unit is assigned a certain predefined number of operating room sessions. Each session has several characteristics: the start and finish time, the doctor or doctors who are going to operate and the specialty or specialties which can be performed in the session. As mentioned in Introduction, the MSS is stable throughout the year.

The waiting list is a dynamic list of patients who have to be operated on. This list is continuously updated. When the tentative schedule is built (2 weeks before the first day of the two-week plan), only those patients on the list at that moment are considered. When the rescheduling is performed, the new state of the list is considered. A surgery may belong either to a specialty of the unit or be classified as "general". A specialised surgery must be performed in a session assigned to the corresponding specialty, while a general surgery can be performed in any session. In addition, we know the estimated duration of the surgery and also if it needs one or two doctors. Furthermore, in some special cases, a surgery needs a specific doctor to perform it. Finally, and due to allergies of some patients, some surgeries have to be the first of their session.

As commented above, we only consider in this problem elective patients. An elective patient is someone for whom the surgery can be planned well in advance. The opposite concept to elective patients would be non-elective or urgent patients, i.e. those who need a surgery urgently. In the hospital of reference, urgent patients are treated elsewhere in the hospital. We work both with inpatients and outpatients. Inpatients must stay at least one night in the hospital, whereas an outpatient enters and leaves the hospital on the same day. The hospital dedicates specific operating rooms for in- and outpatients, and therefore we deal with those patients in the same way.

For each patient, the following information is known: the date of admission, his/her condition of inpatient or outpatient and his/her priority. The priority can be 1, 2 or 3, depending on how imperative the surgery is. See e.g. Agnetis et al. (2012) for other examples of similar definitions. It is important to remark that we are going to deal with durations as constants, not as random variables. The hospital of reference is a training facility, too. In those surgeries that are going to finish early, doctors let trainees do more of the work than in surgeries that are more pressed for time. Furthermore, between surgeries there must be a waiting time for cleaning fixed at 20 min.

Regarding the objective functions, the hospital is interested in two performance criteria. The most important goal is to schedule patients before their (soft) due dates. Patients with priority 1, 2 and 3 should be operated on, according to the local government, in less than 1, 3 and 12 months, respectively. This is a goal and not a restriction, since due dates are not hard constraints. The way to deal with this goal is a key topic in this paper. The way to cope with similar goals in the literature is through the minimisation of waiting times. For the time being, we will call this objective function due dates satisfaction, *DDS*. The other objective function, as already commented on in Introduction, is the utilisation rate of the surgery rooms, *UR*.

2.2 Problem formulation

Let us describe the model. There are n surgeries $\{O_1, O_2, \ldots, O_n\}$ that can be selected to be performed in a two-week plan. There are m sessions $\{S_1, S_2, \ldots, S_m\}$ available. We have to decide which surgeries are going to be performed and in which session. Taking into account the specifications of a surgery (specialty and specific doctor), we can calculate before the scheduling if the surgery can be performed in a certain session. With this information, we create a set of feasible sessions for each surgery, *FeasSes*(O_i). Note that the only condition this set does not take into account is whether the surgery is going to be the first in the session. We use the following notation:

Indexes
$i = 1, \ldots, n$ (number of surgeries to perform)
$j = 1, \ldots, m$ (number of sessions)
Constants
d_i = duration of surgery O_i
D_j = duration of session S_j
dd_i = due date of surgery O_i
wt = waiting time between surgeries
$f_i = 1$ if surgery O_i must be scheduled at the beginning of a session, 0 otherwise
FeasSes(O_i) = set of feasible sessions for surgery O_i
Variables
$x_{ij} = 1$ if surgery O_i is assigned to session S_j, 0 otherwise. The model considered is:

$$Max \ DDS(x) \tag{1}$$

$$Max \ UR(x) = \frac{\sum_{i=1}^{n} d_i \sum_{j=1}^{m} x_{ij}}{\sum_{j=1}^{m} D_j} \tag{2}$$

$$\sum_{j=1}^{m} x_{ij} \leq 1 \qquad (i = 1, \ldots, n) \tag{3}$$

$$\sum_{i=1}^{n} (d_i + wt) \, x_{ij} \leq D_j + wt \qquad (j = 1, \ldots, m) \tag{4}$$

$$\sum_{i=1}^{n} f_i x_{ij} \leq 1 \qquad (j = 1, \ldots, m) \tag{5}$$

$$x_{ij} = 0 \ if \ j \notin FeasSeas(O_i) \tag{6}$$

$$x_{ij} \in \{0, 1\} \quad (i = 1, \ldots, n; j = 1, \ldots, m). \tag{7}$$

Objective function (1) will be discussed in Sect. 2.3. Objective function (2) measures the utilisation rate of the operating rooms. Intentionally, *UR* does not take into account the waiting time, because in this way it is calculated by the local government. Restrictions (3) guarantee that a given patient is scheduled at most once. Inequalities (4) take into account the time availability of a session. The addition of the waiting time on the right-hand side corresponds to the fact that the first surgery of a session does not have a waiting time before it. A session can have at most one surgery that has to be necessarily scheduled at the beginning of that session, as stated in restrictions (5). Restrictions (6) restrict the assignment of

a given surgery only to feasible sessions. These restrictions fix many variables in the model.

2.3 Short-term objective functions

The local government evaluates the hospital according to the percentage of tardy patients and to the utilisation rate of the operating rooms. However, by optimising the *DDS*, we obtain values of the utilisation rate above the desired values of the hospital. The reason is that every objective function defined below favours the inclusion of one patient more if it is possible, independently of his/her due date. With this the utilisation rate is also favoured. The DDS and the utilisation rate are not negatively correlated. Summing up, we will focus on minimising the percentage of tardy patients.

As commented above, a tardy patient is one who is operated on later than his/her due date. The percentage of tardy patients is a long-term goal, which cannot be captured in a two-week programming in a straightforward way. One of the purposes of this paper is to define and compare short-term objective functions in the long term, to see which one leads to the best result according to this goal.

The *DDS* objective functions we are going to propose allocate weights to the assignment of an surgery O_i to a session S_j. Common sense dictates that, to minimise the percentage of tardy patients, these weights should increase as the day of the session approaches the due date of the surgery. However, note that if we only want to minimise this goal in a short-term optimisation problem, surgeries after their due date should simply not be scheduled. However, a hospital cannot follow this policy. Therefore, we have imposed a condition on the objective functions we have designed: the weight allocated to the assignment of a surgery to a session after the due date should be at least as big as the weight just before the due date. In the computational results, we also use the tardiness for tardy patients as a performance measure.

Let us define the objective functions we are going to consider. All objective functions except case C) should be maximised. The comparison among them is done in Sect. 4.2.

(A) **SCORE function** (Agnetis et al. 2012)

We will use an objective function that appears in Agnetis et al. (2012) as a stepping stone for our proposals. According to that paper, the "management objectives are to maximise the utilisation of ORs, without resorting to overtime and, as far as possible, perform each case surgery within the respective due date". Therefore, the paper optimises the same objectives as us. The function is:

$$DDS_1(x) = SCORE(x) = \sum_{i=1}^{n} \sum_{j=1}^{m} K_{ij} x_{ij}$$

being $K_{ij} = d_i(W_i - R_{ij})$, W_i = maximum allowed time in waiting list for surgery O_i, and R_{ij} = slack time (days to due date) of surgery O_i if scheduled in session j. Note that R_{ij} is

negative for tardy surgeries and represents in those cases the tardiness but with negative sign. In the original function, a constant K_i not depending on the session was used; although not stated in the paper, we suppose the reason was that there was only one session for each subspecialty.

(B) **LASTDAY function**

In this function, there are two weights that affect the inclusion of a surgery in a session. The first one helps select those surgeries whose due dates are closer to expire. The second one tries to schedule surgeries as soon as possible inside the planning period.

$$DDS_2(x) = LASTDAY(x) = \sum_{i=1}^{n} (H - dd_i)$$
$$\sum_{j=1}^{m} (lastday - day(j) + 1) x_{ij}.$$

H is a number bigger than all due dates. The value *lastday* represents the last day of the planning period. In our case *lastday* = 14. The value $day(j)$ gives the day of the session $S_j(1, 2, \ldots, lastday)$, depending on the day of the session within the 2 weeks. We add one so that the value is not 0 when we are working with sessions on the last day.

(C) **TARDINESS function (minimisation)**

The tardiness of an operation O_i depends on the session S_j when it is scheduled and it calculates the number of days an operation is tardy if the operation is tardy, and 0 otherwise. Hence, $t_{ij} = \max\{0, -R_{ij}\}$. The function *TARDINESS* sums three quantities. The first one sums the tardiness of scheduled surgeries. The second term only adds for unscheduled operations and uses t'_i, which calculates the tardiness of operation $O(i)$ if it was scheduled the next planning day outside the planning period. The last quantity comes into action when no surgery is tardy. In that case, we apply the objective function of (A):

$$DDS_3(x) = TARDINESS(x) = \sum_{i=1}^{n} \sum_{j=1}^{m} t_{ij} x_{ij}$$
$$+ \sum_{i=1}^{n} t'_i \left(1 - \sum_{j=1}^{m} x_{ij}\right) - \beta SCORE(x).$$

The parameter β is a number small enough so that the third term only serves to untie. We have used $\beta = 1/10,000$.

(D) **LINE function**

To describe the *LINE* function, we need to define additional variables and constants:

Constants

$L_{ij} = 1$ if surgery O_i is tardy and assigned to session S_j, 0 otherwise

nL_{ij} = number of days the surgery O_i is tardy if it is assigned to session S_j, 0 if the surgery is not tardy or it is not assigned to that session

$$maxnL_i = \max_{\{j=1,\ldots,m\}} \{nL_{ij}\}$$

$weight_i$ = weight associated to surgery O_i. It is explained below.

Variables

$z_i = 1$ if surgery O_i is performed in this planning period, and it is performed after its due date, 0 otherwise.

The function LINE is defined as follows:

$$DDS_4(x, z) = LINE(x, z) = \sum_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} weight_i \cdot$$

$$x_{ij} - \sum_{i=1,\ldots,n} z_i + \frac{\sum_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} (maxnL_i - nL_{ij}) x_{ij}}{1 + \sum_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} maxnL_i}.$$

We need to add the following restrictions to the model of Sect. 2.2, where (8) assures that z_i is 1 if surgery O_i is performed in this planning period and is performed after its due date.

$$z_i \geq \sum_{j=1,\ldots,m} L_{ij} x_{ij} \quad (i = 1, \ldots, n) \tag{8}$$

$$z_i \in \{0, 1\} \, \forall i = 1, \ldots, n. \tag{9}$$

The first term in $LINE(x,z)$ selects the surgeries to be included in the schedule. The employment of the weight favours the operations closer to their due date and even more those already tardy. The second term minimises the number of tardy surgeries. The third term minimises the tardiness of tardy surgeries. The first term uses the weights $weight_i$ to select the surgeries. We have grouped surgeries in three sets: (a) tardy surgeries, those which are already tardy, (b) critical surgeries, those which are close to their due dates and (c) non-critical surgeries, those which are still not close to their due dates. The weights of the surgeries in each one of these groups will be calculated with a different linear function based on the slack (R_i), defined as the number of days until the due date of the surgery O_i if it is scheduled in the next planning day outside the planning period. After some preliminary tests (with other random instances different from the ones described in Sect. 4), we have decided to work with the weights described and depicted in Fig. 1.

2.4 Variations in previous functions

We have also studied the effect of two additional elements that can be applied to the previous and other objective functions. The first element is the duration of the surgery we want to select, d_i. The second element has to do with the specialty of the surgery. We should try to promote the selection of surgeries that belong to specialties that have many surgeries or that have few sessions in the week. For this we have defined the Dynamic Specialty Load (*DSL*). It takes into account the
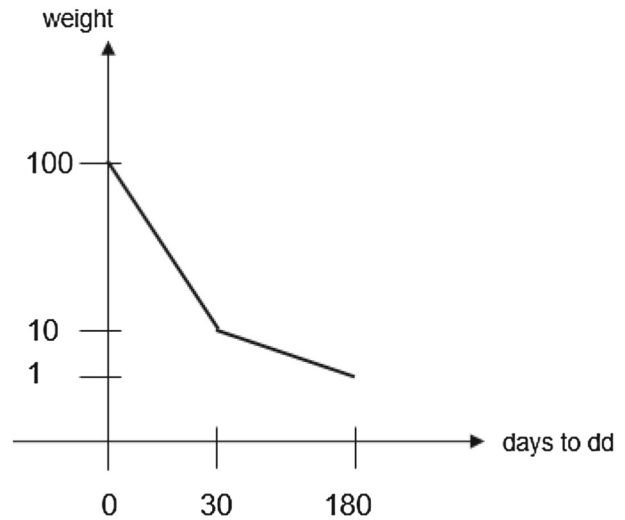

Fig. 1 Weights assigned to the surgeries in LINE

percentage of surgeries in a specialty as a function of the total number of surgeries. This number can be calculated by averaging the historic frequencies of the specialties and the actual frequencies in the current waiting list. We divide this sum by the percentage of sessions of that specialty with respect to the total number of sessions. The value $sp(i)$ denotes the specialty of surgery O_i and "#" denotes number.

$$DSL(i)$$
$$= \frac{0.5 \, (historic \, frequency \, of \, sp(O_i) + actual \, frequency \, of \, sp(O_i))}{\# sessions \, of \, sp(O_i) / m}.$$

Both $DSL(i)$ and d_i are constants that can be multiplied by x_{ij} in the objective functions. We have studied whether it is useful to multiply one of them, both or none.

3 Rescheduling

The tentative schedule for a planning period is calculated several days before the actual period, in order to inform the patients who are going to be operated on, to inform the doctors, to prepare special equipment for some surgeries, etc. In our case, the time gap before the planning period is 2 weeks. However, a few days before the actual period, the schedule is revised, with the new information available. On the one hand, some surgeries should be cancelled: (a) if the patients have an unavoidable appointment elsewhere, or have already been operated on in emergencies due to complications of their illnesses, etc.; (b) if the doctors are unable to perform their surgeries; (c) if other essential resources such as special equipment are unavailable. On the other hand, there are new patients on the waiting list who could be included in the final schedule. This revision of the initial schedule can lead to cal-

culating a new and final schedule. This process is called the rescheduling phase or rescheduling.

Rescheduling elective patients presents several disadvantages. The most important one is the inconvenience to the patients provoked by the changes: shifting a surgery one or several days, cancellations of scheduled surgeries or the incorporation of new surgeries into the final plan. Moreover, there are also inconveniences forced on doctors who may have to perform surgeries different than planned. Nevertheless, the benefits of rescheduling may be significant: a larger number of patients may be operated on, a better utilisation of the OR could be achieved and, in general, better values in the hospital goals may be obtained. The hospital should decide to what extent to change the preliminary schedule. The method presented in this paper to calculate this final schedule allows the hospital to take this decision while controlling the changes.

As we have seen in Introduction, there are not many papers in the literature that consider rescheduling in hospitals. Almost all of these papers consider a different problem: the schedule created for elective patients have to make room for non-elective (urgent or emergency) patients. It is worthwhile mentioning that this section is useful for any hospital that reschedules elective patients.

The tentative schedule, $S_{tentative}$, is obtained by solving the model defined in Sect. 2.2. This schedule is obtained two weeks before the planning period with the information available at that time. Let us call f the objective function used in this problem (any function defined in Sect. 2.3 or any other objective function). We assume the problem is a single objective optimisation problem. We call f_{MAX} the value of f when optimising the same model but with the information available in the rescheduling phase, in our case one week before the planning period. When obtaining the final solution, the hospital wants to optimise the same objective function f, but does not want to obtain a plan with many changes with respect to $S_{tentative}$. We define a method to obtain the final schedule that reflects the changes the hospital would be willing to allow. We define a *policy of changes P* as a set of rules imposed by the hospital that determines which changes can be made and which changes are not allowed when building the final schedule. Hence, many feasible schedules are forbidden, because they do not comply with the policy of changes. Let us define $S(P)$ as the set of schedules fulfilling the rules imposed by policy P. We propose the following general method to calculate the final schedule:

Max $f(S)$

s.t. Restrictions of model in Sect. 2.2.

$S \in S(P)$

We have defined four different policies. Each one is applied solving the model in Sect. 2.2 with the additional constraints indicated in each case. Section 4.3. compares these policies. Policies use the concepts of problematic and unproblematic patients. We define a problematic patient as a patient who is not available to be operated on in the session he/she was planned in $S_{tentative}$, due to, for example, any of the reasons presented in the first paragraph of this section. An unproblematic patient is a scheduled patient who can still be operated on in the planned session. Let us define the policies:

- **Policy 1** Unproblematic patients must be operated on in the same session

We define:

UnP_1 = {unproblematic patients scheduled in the first week}.

UnP_2 = {unproblematic patients scheduled in the second week}.

$S(i)$ = session assigned to patient O_i, if any; $\emptyset$ if the patient is not scheduled.

We add the following to the formulation of Sect. 2.2:

$$x_{ij} = 1 \ (i \in UnP_1 \cup UnP_2; \ j = S(i)).$$

- **Policy 2** Unproblematic patients scheduled in the first week must be operated on in the same session. Unproblematic patients scheduled in the second week must be operated on, independently of the session.

We define:

We add the following to the formulation of Sect. 2.2:

$$x_{ij} = 1 \quad (i \in UnP_1; \ j = S(i))$$
$$\sum_{j=1}^{m} x_{ij} = 1 \quad (i \in UnP_2).$$

- **Policy 3** Unproblematic patients must be operated on, independently of the session.

We add the following to the formulation of Sect. 2.2:

$$\sum_{j=1}^{m} x_{ij} = 1 \quad (i \in UnP_1 \cup UnP_2).$$

- **Policy 4** The same as Policy 3, but only for patients with priorities 1 or 2. Surgeries of patients with priority 3 may be cancelled.

We define:

Pr_1 = {patients with priority 1}, Pr_2 = {patients with priority 2}.

Table 1 Contributions to function $dist(S)$

Contributions of scheduled patients	Contributions of unproblematic patients	0	If the patient is operated on in S in the same session as in $S_{tentative}$
		5	If the patient is operated on in S in the same week as in $S_{tentative}$ but in a different session
		10	If the patient is operated on in S in a different week than in $S_{tentative}$
		20	If the patient is not operated on in S
	Contributions of problematic patients	1	If the patient is operated on in S in a different week than in $S_{tentative}$
		2	If the patient is not operated on in S
Contributions of not- scheduled patients		0	If the patient is not operated on in S
		1	If the patient is operated on in S in the second week
		3	If the patient is operated on in S in the first week

We add the following to the formulation of Sect. 2.2:

$$\sum_{j=1}^{m} x_{ij} = 1 \quad (i \in (UnP_1 \cup UnP_2) \cap (Pr_1 \cup Pr_2)).$$

A different goal can be followed when calculating the final schedule once the new information is available. It consists in obtaining a schedule as similar as possible to the initial schedule $S_{tentative}$. We define $S_{minDist}$ as the final schedule that provokes the minimum inconveniences to the hospital with respect to $S_{tentative}$. In order to calculate this schedule, we introduce a distance function, $dist$, which calculates the differences between a schedule S and the tentative solution. Each hospital should define therefore their own function $dist$. This function $dist$ is defined according to the preferences of the specific hospital. This function should measure the inconveniences caused to the doctors, hospital and patients when implementing a schedule S instead of continuing with the initial schedule $S_{tentative}$. We have worked with a linear distance function that sums the contribution of scheduled patients and the contribution of not-scheduled patients. A scheduled patient is a patient who was going to be operated on in $S_{tentative}$. The contribution of scheduled patients is the sum of the contributions of problematic and unproblematic patients.

The contributions of patients to $dist(S)$ are detailed in Table 1. We have considered several assumptions due to the specific characteristics of the hospital of reference and the two-week plan. Although the cancellation of a surgery can be due to patient or hospital causes, we only consider the first ones. Furthermore, we have assumed that there are just two types of problematic patients, those who cannot be operated on in the week he/she was planned and those who cannot be operated on in the whole two-week period. At any rate, meth-

ods presented do not depend on the contributions considered and their specific values.

Once this distance is defined, we can calculate the schedule $S_{minDist}$ that minimises the distance. Let us call $f_{MIN} = f(S_{minDist})$. To calculate $S_{minDist}$, we use the model of Sect. 2.2 replacing the objective function by $dist(S)$ and updating the set $FeasSes(O_i)$, to prevent patients being scheduled when they are not available. We use schedule $S_{minDist}$ as a keystone to analyse the benefits of allowing changes.

4 Computational results

In this section, we present computational results for the methods introduced in the two previous sections. To assess the efficiency of the proposed methods, the computational experiments have been carried out on realistic instances, randomly generated following the data from a hospital. These instances are described in Sect. 4.1. Section 4.2 compares the results obtained by the objective functions proposed in Sect. 2.3. Section 4.3 handles the rescheduling problem treated in Sect. 3.

4.1 Random instances

Using historic information provided by the hospital, we have generated several real-life-based data waiting lists for a pre-established initial date. Their main characteristics are:

- The value n is randomly selected in the interval [135,189].
- Surgeries are assigned a random priority in such a way that 1% have priority 1, 34% have 2 and 65% have 3.

Table 2 Results for 1% or 1-min limitation

	UR	%Tardy	Avg_Tard_Tardy	#Operated	Total_Tard_Operated	#Tardy_Wait	Total_Tard_Wait
SCORE	89.32	11.20	14.47	1715.90	3105.05	82.05	2334.90
LINE	87.34	6.38	13.12	1737.78	1484.43	47.75	1767.90
LINE_D	88.37	7.95	17.26	1710.93	2579.63	75.40	2190.03
LINE_L	87.29	5.73	11.30	1731.20	1159.10	47.35	1944.90
LINE_DL	88.24	6.78	12.94	1708.28	1761.58	69.23	2276.88
TARDINESS	88.57	5.88	5.54	1755.05	551.65	40.70	1953.00
TARDINESS_D	88.69	7.52	11.37	1713.93	1656.13	77.58	3004.13
TARDINESS_L	88.63	5.62	6.06	1753.05	601.00	43.73	2142.03
TARDINESS_DL	88.73	7.30	9.88	1714.80	1444.73	72.73	3004.58
LASTDAY	87.07	6.26	29.97	1845.18	4013.20	89.75	5836.18
LASTDAY_D	88.29	10.55	22.63	1709.48	4547.68	111.48	4182.63
LASTDAY_L	87.08	6.58	30.22	1838.25	4097.50	87.58	6079.08
LASTDAY_DL	88.07	9.81	24.46	1710.93	4166.30	105.80	5856.90

- 2% of surgeries must be scheduled at the beginning of the session.
- The specialty of a surgery is probabilistically chosen based on the percentage of surgeries of that specialty in the data provided by the hospital.
- The duration of a surgery is determined probabilistically based on the number of surgeries along the year that have their duration in certain pre-established intervals. The intervals we have considered are of 15 min, starting at 0 and finishing after 6 h.

The characteristics of the sessions are fixed and follow the values used in the hospital. There are 8 sessions per week: one session of 12 h and the rest of seven. Each session has one or two specialties assigned, and "general surgeries" can be assigned to all sessions.

To incorporate more diversity to the instance set, we have doubled the number of patients in the initial waiting list, the number of new patients per week and the number of sessions. Therefore, we consider *ProblemSize* = 1, 2. We have also considered two levels for the incorporation of new patients each week: the "normal level" and a "congested level". The congested level incorporates a 25% more new patients each week. The normal level is defined by *Congestion* = 1, and the congested level is defined by *Congestion* = 1.25. For each combination of the possible options (*ProblemSize* × *Congestion* = 2 × 2 = 4), we have randomly generated 10 instances, making a total of 40 instances. To increase the precision of our comparisons, we use common random numbers.

To study the long-term behaviour, we follow the papers of Persson and Persson (2009) and Agnetis et al. (2012). We successively perform two-week plans for each instance, starting from the pre-established initial date and until a year has passed (i.e. a total of 26 time plans for each instance). The waiting lists are updated after each two-week plan, removing the surgeries already performed and incorporating new random patients following the data provided by the hospital.

Tests have been carried out on an Intel(R) Core(TM) i7-2670QM CPU 2.20 GHz 6 GB RAM. Models have been solved with GAMS 23.3, using CPLEX solver.

4.2 Comparing short-term objective functions

Following Sect. 2.4, we have considered four possibilities for the three objective functions we have defined: (a) the function itself, (b) the function where x_{ij} is multiplied by the duration, named _D, (c) the function where x_{ij} is multiplied by the *DSL*, adding _L to the name, and (d) another where x_{ij} is multiplied by the duration and the *DSL*, named _DL. To compare the functions, we have run GAMS with two different limits. In Table 2, the algorithm stops after 1 min of time per optimisation problem or when the gap with the lower bound reaches 1%. In Table 3, the algorithm cuts the search after 2 min. To give an example of the gap to the lower bound, the average deviation with respect to the bound provided by GAMS is less than 1% in the cases of 2 min of SCORE and LINE_L (0.962 and 0.085%, respectively).

We have run the models for one year and calculated the averages over all weeks except the first 2 weeks. We have decided to leave out the first plan because there were already many tardy patients and some of them were many days tardy. Including these patients would disturb the results, as no function can schedule them before their due dates. Furthermore, their big values in tardiness disturb the averages of these values for the rest of patients.

We have used several performance measures to compare the objective functions. The most important one is the percentage of tardy patients (*%Tardy*). However, the average

Table 3 Results for 2-min limitation

	UR	%Tardy	Avg_Tard_Tardy	#Operated	Total_Tard_Operated	#Tardy_Wait	Total_Tard_Wait
SCORE	89.39	11.11	14.35	1718.28	3055.13	81.00	2300.83
LINE	88.80	5.23	12.05	1764.55	1162.90	41.45	1518.25
LINE_D	89.19	6.88	15.85	1731.30	2165.43	67.93	1893.13
LINE_L	88.79	4.61	10.60	1760.75	928.40	41.85	1677.43
LINE_DL	89.18	6.03	12.05	1729.60	1545.50	62.93	1996.80
TARDINESS	89.03	5.70	5.81	1757.73	546.53	38.10	1791.38
TARDINESS_D	89.33	7.05	11.34	1719.33	1486.80	75.13	2878.05
TARDINESS_L	89.05	5.47	5.82	1754.08	523.33	40.80	2008.65
TARDINESS_DL	89.27	6.93	9.57	1723.93	1242.35	72.08	3033.95
LASTDAY	87.35	6.21	29.94	1848.40	3974.15	88.63	5783.28
LASTDAY_D	88.41	10.44	23.21	1712.45	4481.48	110.60	4117.05
LASTDAY_L	87.41	6.46	29.89	1843.53	4001.30	86.78	6032.90
LASTDAY_DL	88.26	9.67	24.72	1714.15	4129.10	104.00	5735.88

tardiness of tardy patients (*Avg_Tard_Tardy*) is also important. Other measures we have incorporated in the table are the number of surgeries in the schedule (*#Operated*), the total tardiness of the operated patients (*Total_Tard_Operated*), the number of already tardy patients on the list after one year (*#Tardy_Wait*) and the total tardiness of those patients (*Total_Tard_Wait*).

The utilisation rate (*UR*) is also a very important goal for the hospital. However, a percentage of 80% is more than enough for the hospital, as happens in all cases of the table (almost 90%). This is one of the reasons why we have focused the paper on the percentage of tardy patients. Considering just the percentage of tardy patients, all our proposals beat the SCORE function. The best functions cut this percentage in half (from 11 to 5–6%). This confirms the importance of choosing the correct objective function. Not every "similar" function is going to lead to the same results.

Regarding the average of tardiness in tardy patients (*Avg_Tard_Tardy*), variations of the LASTDAY function obtain poor results in this field, multiplying by two the average days with respect to the SCORE function. Variations of LINE obtain, in general, similar results to SCORE, and the functions belonging to TARDINESS improve the results of SCORE. Taking into account the performances *%Tardy* and *Avg_Tard_Tardy*, the best four functions are TARDINESS, TARDINESS_L, LINE (especially in Table 3) and LINE_L. LINE_L outperforms LINE in all measures except in the *UR* and *#Operated*, which are very similar. Another proof that these functions handle the waiting list better comes from the rest of the performances in the tables. They occupy the first 4–5 positions in every performance, except the number of surgeries. For instance, they clearly beat the rest of the functions in having a fewer number of tardy patients, both in Tables 2 and 3.

If we compare the results of Tables 2 and 3, we can see that there are functions which improve their results, whereas others stay almost the same (SCORE and variations corresponding to LASTDAY and TARDINESS). The functions that most improve their results are those corresponding to LINE, especially in *%Tardy*. Note that a decrease in this percentage makes it more difficult in general to reduce the average tardiness of tardy patients. The reason is that the patients that have been transformed from tardy patients to patients "on time" are usually those patients with very few days tardy. The average of the rest of the tardy patients can therefore be easily higher.

Regarding the inclusion of the duration and *DSL*, the duration always worsens *%Tardy*. The *DSL* improves this measure in LINE and TARDINESS, but not in LASTDAY. At any rate, *DSL* is clearly a parameter to consider when designing an objective function for this goal.

4.3 Results for the rescheduling problem

Table 4 shows the results for the four rescheduling policies defined in Sect. 3. We have assumed a percentage of problematic patients of 10%: 5% who cannot be operated on in the same week and 5% who cannot be operated on at all in the 2 weeks. The function used to calculate the tentative schedule before optimising the rescheduling problem has been LINE_L. We cut the search after 2 min, and the average deviation with respect to the bound provided by GAMS is 0.06%. To compare the results of the policies, we have also included the results of $S_{minDist}$, the final schedule that provokes the minimum inconveniences to the hospital with respect to $S_{tentative}$. The meanings of the columns are as follows. The percentage *%Deviation* calculates the deviation with respect to f_{MAX}. As stated in

Table 4 Results for the rescheduling problem

	%Deviation	#Cancelled_Surgeries (unproblematic/all)	#New_Surgeries	#Changed_Sameweek	#Changed_ Difweek
$S_{minDist}$	12.35%	0/6.78	0	0	2.43
Policy 1	7.92%	0/8.08	8.83	0	1.13
Policy 2	1.34%	0/7.93	7.55	5.18	3.88
Policy 3	0.84%	0/7.58	6.98	7.25	4.73
Policy 4	0.01%	26.25/32.48	23.40	10.03	7.33

Sect. 3, f_{MAX} is the value of f when optimising the model with the information available in the rescheduling phase. The two numbers in #Cancelled_Surgeries (*unproblematic/all*) show the average number of surgeries that are cancelled from $S_{tentative}$ to the new schedule, taking into account only unproblematic patients and all patients, respectively. Note that #Cancelled_Surgeries (*all*) include patients who cannot be scheduled because of their unavailability in the planning period. The other three columns are self-explanatory.

As it was expected, the more changes we allow in the schedule, the better the objective function. That is, a bigger flexibility in the schedule would allow better results in terms of *%Tardy* and other goals. The difference between implementing the schedule with minimum distance and the schedule of f_{MAX}, which corresponds to allowing total flexibility, is more than 12%. Policy 1 improves almost 4.5% by eliminating on average 8 surgeries from problematic patients, changing 1 problematic patient to a different week and introducing nearly 9 new surgeries. Policy 4 performs a huge number of changes, whereas Policy 3 seems to be the best option in terms of the number of changes and improvement in the objective function.

Another aspect not appearing in the table is worth mentioning. The gap between $S_{minDist}$ and f_{MAX} changes drastically from instance to instance. In 37.5% of the instances, the gap is less than 1%, whereas in 27.5% the gap is more than 20%.

5 Conclusions

In this paper, we have studied the problem of creating a schedule of elective surgery patients for a unit of a hospital based on a real problem. Two phases are followed. First, a tentative schedule is created 2 weeks before the schedule starts. Second, a final plan is created a few days before, rescheduling the tentative schedule with the new information available.

In the first part of the paper, we have proposed several objective functions trying to capture the goal of minimising the percentage of tardy patients, without forgetting the average tardiness of these patients. The best of our proposals halves the percentage of tardy patients of our keystone,

one of the few functions in the literature explicitly used for this goal. At the same time, some of these functions lower the average tardiness of tardy patients. At any rate, our study clearly proves that short-term objective functions that look for the same long-term goal may offer great differences in that goal. Therefore, any hospital with a long-term goal should perform a similar study to the one provided here.

In the second part, we have described a general method to reschedule elective patients: optimising the original objective function while restricting the changes in the original plan to a policy of changes defined by the hospital. It is worthwhile to mention that the rescheduling method presented can be adapted in a straightforward way to policies of changes different to the ones considered in this paper.

We have seen in our computational results that there are instances where the possible improvement is low, and therefore a policy of little disturbance should be adopted. However, there are instances where a very important gain can be achieved. A greater flexibility from the hospital can lead to a substantial improvement in its goals.

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), 206–232.

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

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

Conforti, D., Guerriero, F., Guido, R., & Veltri, M. (2011). An optimal decision-making approach for the management of radiotherapy patients. *OR Spectrum*, *33*, 123–148.

Conway, R., Maxwell, W., & Miller, L. (1967). *Theory of scheduling*. Reading, MA: Addison-Wesley.

Demeulemeester, E., Beliën, J., Cardoen, B., & Samudra, M. (2013). Operating Room Planning and Scheduling, Chapter 5 (pp. 121-

152) in Handbook of Healthcare Operations Management (Methods and Applications), Brian T. Denton (ed).

Erdem, E., Qu, X., & Shi, J. (2012). Rescheduling of elective patients upon the arrival of emergency patients. *Decision Support Systems*, *54*, 551–563.

Fei, H., Meskens, N., & Chu, C. (2010). A planning and scheduling problem for an operating theatre using an open scheduling strategy. *Computers & Industrial Engineering*, *58*, 221–230.

Guerriero, F., & Guido, R. (2011). Operational research in the management of the operating theatre: a survey. *Health Care Management Science*, *14*(1), 89–114.

Hanset, A. (2010). Amélioration de l'efficacité de la programmation opératoire au sein des établissements de soins de santé, Ph.d. Thesis, December 10, 2010, Catholic University of Mons (FUCAM), Belgium.

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

Li, Q. (2010). Multi-objective Operating Room Planning and Scheduling, Ph.D. Dissertation.

Li, F., Gupta, D., & Potthoff, S. (2013). Improving Operating Room Schedules (November 20, 2013). SSRN: https://doi.org/10.2139/ssrn.2357535.

Luo, L., Luo, Y., You, Y., Cheng, Y., Shi, Y., & Gong, R. (2016). A MIP model for rolling horizon surgery scheduling. *Journal of Medical Systems*, *40*, 127.

Macario, A., Vitez, T. S., Dunn, B., & McDonald, T. (1995). Where are the costs in perioperative care? Analysis of hospital costs and charges for inpatient surgical care. *Anesthesiology*, *83*(6), 2–4.

Meskens, N., Duvivier, D., & Hanset, A. (2013). Multi-objective operating room scheduling considering desiderata of the surgical team. *Decision Support Systems*, *55*, 650–659.

Persson, M., & Persson, J. A. (2009). Health economic modeling to support surgery management at a Swedish hospital. *Omega-International Journal Management Science*, *37*(4), 853–863.

Reilly, T. A., Marathe, V. P., & Fries, B. E. (1978). A delay-scheduling model for patients using a walk-in clinic. *Journal of Medical Systems*, *2*(4), 303–313.

ValiSiar, M. M., Gholami, S., & Ramezanian, R. (2017). Multi-period and multi-resource operating room scheduling and rescheduling using a rolling horizon approach: A case study. *Journal of Industrial and Systems Engineering.*, *10*, 97–115.

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 Management Science*, *15*(4), 355–372.

van Essen, J. T., Hurink, J. L., & Hartholt, W. (2013). Rescheduling in the OR: A Decision Support System Keeps Stakeholders and OR-Manager Happy. *Operating Room Supplement*, *15*(4),

Velasquez, R., Melo, T., & Kufer, K. H. (2008). Tactical operating theatre scheduling: Efficient appointment assignment. *Operations Research Proceedings*, *2007*, 303–308.