Column generation approach to operating theater planning with elective and emergency patients

Abstract

The elective surgery planning problem for operating rooms shared between elective and emergency patients is addressed. The planning problem consists in determining the set of elective patients to be operated on in each operating room in each period over a planning horizon in order to minimize patient-related costs and the expected operating rooms' utilization costs. A stochastic mathematical programming model and a column generation approach are proposed. The proposed approach results in both a near-optimal solution and a lower bound to assess the degree of optimality. Solutions within 2% of the optimum are obtained in a short computation time for problems of practical sizes with 12 operating rooms and about 210 elective patients.

Keywords: Operating theater planning, stochastic model, column generation, emergency surgery

1. Introduction

A hospital operating theater generally admits two categories of patients: elective patients, who can be planned in advance and emergency patients, who arrive randomly during the day and must be served urgently. One of the major problems facing the operating theater manager is how to allocate the capacity of the Operating Rooms (ORs) among these two competing groups of patients. Emergency surgery demand arises randomly almost every day and must be met on the same day because of the critical condition of emergency patients. Devoting a large amount of the capacity to emergency surgery will increase OR idle times and the waiting list for elective surgery. However, if the capacity reserved for emergency surgeries is insufficient, this may result in excessive overtime usage and cancellation or postponement of already planned elective patients; which can significantly degrade the quality of service, and generate additional costs and a bad image for the hospital.

This paper addresses the problem of elective surgery planning under uncertain demand for emergency surgery. The problem consists in determining a plan that specifies the set of elective patients that would be operated on in each OR in each period over a planning horizon. The surgery plan should minimize costs related to performing elective surgeries and the expected utilization costs of the ORs, i.e., expected overutilization and underutilization costs.

ORs are among the most critical and expensive resources at most hospitals. For this reason, the planning and scheduling of ORs has been widely addressed in the healthcare literature. Belien and Demeulemeester (2007) developed analytical models and solution heuristics for building cyclic master surgery schedules to minimize the expected total bed shortage. Vissers *et al.* (2005) Vissers, J. M. H., Adan, I. J. B. F. and Bekkers, J. A. 2005. Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management Mathematics*, 16: 281–

304. [CrossRef], [Google Scholar] studied the problem of creating a master surgery schedule while considering multiple resources; they formulated the problem as a mixed-integer program and solved it heuristically. Marcon *et al.* (2003) Marcon, E., Kharraja, S. and Simmonet, G. 2003. The operating theatre planning by the follow-up of the risk of no realization. *International Journal of Production Economics*, 85: 83–90. [Google Scholar] proposed a tool to assist the planning negotiation between the different actors of the surgical suite by estimating the risk of non-realization of tentative surgery plans. A two-step approach was proposed by Guinet and Chaabane (2003) Guinet, A. and Chaabane, S. 2003. Operating theatre planning. *International Journal of Production Economics*, 85: 69–81. [CrossRef], [Web of Science ®], [Google Scholar] and Jebali *et al.* (2006) Jebali, A., Hadjalouane, A. B. and Ladet, P. 2006. Operating rooms scheduling. *International Journal of Production Economics*, 99: 52–62. [CrossRef], [Web of Science ®], [Google Scholar] for the operating theater scheduling problem. First, elective patients are assigned to ORs. Patients assigned to each OR are then scheduled on a daily basis in order to satisfy material and personnel-related constraints. Integer programming models were developed for the patient assignment problem. Fei *et al.* (2008) Fei, H., Chu, C., Meskens, M. and Artiba, A. 2008. Solving surgical cases assignment problem by a Branch-and-Price approach. *International Journal of Production Economics*, 112: 96–108. [CrossRef], [Web of Science ®], [Google Scholar] used a branch-and-price algorithm to solve the elective surgery planning while minimizing OR overutilization and underutilization. Hans *et al.* (2008) Hans, E.-W., Wullink, G., Houdenhoven, M.-V. and Kazemier, G. 2008. Robust surgery loading. *European Journal of Operational Research*, 185: 1038–1050. [CrossRef], [Web of Science ®], [Google Scholar] considered the problem of assigning elective surgeries and planned capacity slack to OR days. The planned slack aims at minimizing overtime by absorbing the variability in the duration of surgical operations. Various heuristic rules were used to solve this problem. Several other approaches to OR planning have been proposed in the literature (Magerlein and Martin, 1978 Magerlein, J. M. and Martin, J. B. 1978. Surgical demand scheduling: a review. *Health Service Research*, 13: 418–433. [PubMed], [Web of Science ®], [Google Scholar]; Dexter, Macario, Traub, Hopwood, and Lubarsky, 1999; Ozkarahan, 2000).

Most existing approaches for OR planning assume that the OR capacity is devoted to a single class of patients: elective patients. Gerchak *et al.* (1996) Gerchak, Y., Diwakar, G. and Mordechai, H. 1996. Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science*, 42: 321–334. [CrossRef], [Web of Science ®], [Google Scholar] addressed the problem of reservation planning for elective patients when the capacity is used to satisfy elective and emergency surgery demands. The focus of their work is on the characterization of the optimal policy that determines, at the beginning of each day, how many additional requests for elective surgery to assign for that day. Though similar to our problem, their model is mono-period with aggregated capacity and does not specify the intervention date and the OR for each elective case.

This work addresses the planning of elective patients at a hospital operating theater over a planning horizon. The operating theater is composed of several ORs and provides service to elective and emergency patients. Elective patients can be planned in advance with a patient-related cost that depends on the OR and on the date of surgery. A random amount of each OR's capacity is used to serve emergency patients. The planning problem consists in assigning elective patients to ORs with the best trade-off between patient-related costs and the expected utilization costs of the ORs.

The planning problem is first modeled as a stochastic integer program. A column-oriented formulation in which each column represents a possible assignment of elective patients to a particular OR in a particular time period is then derived. The linear relaxation of the latter formulation is then solved via column generation, where the pricing problem, a stochastic knapsack problem, is solved by a dynamic programming method. A feasible plan is derived from the solution of the relaxed problem by a heuristic, and improved by using local optimization heuristics.

This work is an extension of our earlier work (Lamiri *et al.*, 2008 Lamiri, M., Xie, X., Dolgui, A. and Grimaud, F. 2008. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, 185: 1026–1037. [CrossRef], [Web of Science ®], [Google Scholar]) in which a stochastic programming model and a Monte Carlo optimization approach were proposed for elective patient planning under uncertain demand for emergency surgery. In this earlier work, only the total operating theater capacity is taken into account and the assignment of patients to ORs is not considered. As a result, the resulting planning might not be feasible. Furthermore, the Monte Carlo optimization approach relies on mixed-integer programming and hence cannot be used to solve problems of realistic size.

The new contributions of this paper can be summarized as follows: (i) a stochastic model for operating theater planning taking into account the random emergency surgery demand, assignment of patients to ORs, more realistic cost structure of OR capacity utilization; and (ii) an efficient column generation approach able to solve realistic size problems with a very small optimality gap.

The remainder of the paper is organized as follows. Section 2 gives a stochastic integer programming formulation of the planning problem. Section 3 presents the column generation approach, the pricing problem solution method, and the construction and improvement of a feasible plan. Numerical results are presented and discussed in Section 4. Section 5 concludes the paper and discusses possible extensions of this work.

2. Problem statement

This paper considers the planning of elective surgery at a hospital operating theater over a finite horizon of H periods (days). The operating theater is composed of S ORs and provides service to two groups of patients: elective patients, that can be planned in advance; and emergency patients, that must be served on the day of arrival.

At the beginning of the horizon, there are a set of N requests for elective surgery. The planning problem consists in determining the set of elective patients to be operated on in each OR and in each time period over the planning horizon. The starting times and sequence of surgical cases for a given OR can be determined at a later stage on a period-to-period basis (Weiss, 1990 Weiss, E. N. 1990. Models for determining started start times and case orderings in hospital operating rooms. *IIE Transactions*, 22: 143–150. [Taylor & Francis Online], [Web of Science ®], [Google Scholar]; Liu and Liu, 1998 Liu, L. and Liu, X. 1998. Dynamic and static job allocation for multi-server systems. *IIE Transactions*, 30: 845–854. [Taylor & Francis Online], [Web of Science ®], [Google Scholar]; Denton and Gupta, 2003 Denton, B. and Gupta, D. 2003. A sequential bounding approach for optimal appointment scheduling. *IIE Transactions*, 35: 1003–1016. [Taylor & Francis Online], [Web of Science ®], [Google Scholar]). In the rest of the paper, we call an OR $s \in \{1, \ldots, S\}$ in a given period (day) $t \in \{1, \ldots, H\}$ an "OR-day" (s, t). The following notations are used throughout the paper:

H	planning horizon;
$t \in (1, \ldots, H)$	time period index;
S	number of ORs;
$s \in (1, \ldots, S)$	operating room index;
N	number of elective cases;
$i \in (1, \ldots, N)$	elective case index;
d_i	time needed to perform elective case i;
e_i	earliest period to perform elective case i;
a_{its}	cost of performing elective case i in OR-day (s, t);
W_{ts}	random variable representing the capacity used by emergency cases in OR-day (s, t);
$g_{ts}(.)$	operating room utilization cost.

In the following, the notations defining the operating theater planning problem and the underlying assumptions are explained and discussed in detail.

2.1. Operating time

Each elective case i has an operating time d_i, which includes surgery duration, pre-surgery set up time and post-surgery OR cleaning time. Operating times can be efficiently estimated by using historical information and/or surgeons' and operating theater managers' expertise (Shukla *et al.*, 1990 Shukla, R. K., Ketcham, J. S. and Ozcan, Y. A. 1990. Comparison of subjective versus data-base approaches for improving efficiency of operating room scheduling. *Health Services Management Research*, 3: 74–81. [PubMed], [Google Scholar]; Wright *et al.*, 1996 Wright, I. H., Kooperberg, C., Bonar, B. A. and Bashein, G. 1996. Statistical modeling to predict elective surgery time: comparison with a computer scheduling system and surgeon-provided estimates. *Anesthesiology*, 85: 1234–1245. [Google Scholar]; Dexter, Traub and Qian, 1999 Dexter, F., Macario, A., Traub, R. D., Hopwood, M. and Lubarsky, D. A. 1999. An operating room scheduling strategy to maximize the use of operating room block time: computer simulation of patient scheduling and survey of patients' preferences for surgical waiting time. *Anesthesia & Analgesia*, 89: 7–20. [PubMed], [Web of Science ®], [CSA], [Google Scholar]).

Existing statistical studies (Zhou and Dexter, 1998 Zhou, J. and Dexter, F. 1998. Method to assist in the scheduling of add-on surgical cases, upper prediction bounds for surgical case durations based on the log-normal distribution. *Anesthesiology*, 89: 1228–1232. [Google Scholar]; Strum *et al.*, 2000 Strum, D. P., May, J. H. and Vargas, L. G. 2000. Modeling the uncertainty of surgical procedure times: comparison of log-normal and normal models. *Anesthesiology*, 92: 1160–1167. [CrossRef], [PubMed], [Web of Science ®], [CSA], [Google Scholar]) show that surgery duration is in general log-normally distributed. Uncertainty related to surgery duration may have a significant impact on the quality of a surgery plan. Explicit modeling of random surgery duration for operating theater planning is an important issue but it is beyond the scope of this paper.

- *Assumption 1*: Operating times d_i are given constants and multiples of some basic time period θ (e.g., θ= 5 minutes). Furthermore, the operating times are assumed to be independent of the OR-days to which elective cases are assigned.
- *Remark 1*: In practice, the operating time of a surgery may depend on the OR to use, since the pre-surgery set up time varies according to the availability of specialist medical equipment. The column generation approach of this paper applies to the case of OR-dependent operating times. However, due to lack of realistic estimation of operating times for different ORs and the relative minor difference in operating times for different ORs in general, we limit ourselves to OR-independent operating times.

2.2. Patient assignment costs

Each elective patient i is characterized by a release date and a set of assignment costs explained in the following. The release period e_i is the earliest period in which the surgery for case i can be performed. It may represent the hospitalization date, date of medical test delivery, etc. Starting from the earliest date, an elective patient can be planned in any period and incurs an assignment cost that depends on the surgery date and on the OR where the operation will be performed. Let a_{its} represent the cost of performing elective case i in the OR-day (s, t) for $t \in \{e_i, \ldots, H, H+1\}$ and $s \in \{1, \ldots, S\}$. A “fictitious” period H+1 is added to the planning horizon in order to gather elective patients that overflow from the current planning and need to be considered in the next horizon. Thus, $a_{i(H+1)s}$ is the cost of excluding case i from the current planning horizon. Obviously, this cost is independent of the OR, i.e., $a_{i(H+1)s} = a_{i(H+1)s'}$ for any $s, s' \in \{1, \ldots, S\}$, and it can be simply denoted by $a_{i(H+1)}$.

By using this cost structure several constraints can be modeled. The cost matrix $[a_{its}]$ can take into account hospitalization costs (Guinet and Chaabane, 2003 Guinet, A. and Chaabane, S. 2003. Operating theatre planning. *International Journal of Production Economics*, 85: 69–81. [CrossRef], [Web of Science ®], [Google Scholar]; Jebali *et al.*, 2006 Jebali, A., Hadjalouane, A. B. and Ladet, P. 2006. Operating rooms scheduling. *International Journal of Production Economics*, 99: 52–62. [CrossRef], [Web of Science ®], [Google Scholar]), penalties for waiting to get onto the schedule (Gerchak *et al.*, 1996 Gerchak, Y., Diwakar, G. and Mordechai, H. 1996. Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science*, 42: 321–

334. [CrossRef], [Web of Science ®], [Google Scholar]), optimal surgery date, patients' or surgeons' preferences, deadlines and OR specificities and availabilities.

If an elective case i is hospitalized in period e_i and waiting for surgery, then cost a_{its} will increase with the delay, and the increasing rate from one period to the next period represents the hospitalization cost per day. If in addition the patient cannot be operated on in some specific period t' due to the unavailability of the surgeon for example, then a large cost is chosen for $a_{it's}$. Furthermore, if case i can be assigned to any OR, then the costs a_{its} will be independent of the OR, i.e., $a_{its} = a_{its'}$ for any $s, s' \in \{1, \ldots, S\}$. However, if case i can be assigned only to a given set of ORs, because some specific medical equipment is needed, the cost of performing the case in an unsuitable OR will be set very large. If case i must be performed before a given period L_i, this constraint can be taken in account by choosing large costs a_{its} for $t > L_i$. Many other situations can also be modeled by adjusting these costs.

- *Remark 2*: Although many hard constraints on the possibility of performing a surgery on a particular OR-day such as the availability of the related surgeon are represented here through large penalty costs, these constraints can actually be explicitly taken into account by considering for each elective case a set of possible OR-days and the column generation approach of this paper applies with some straightforward changes. This is not introduced in this paper for simplicity of notation.
- *Remark 3*: As explained above, the cost matrix $[a_{its}]$ summarizes different factors to be considered in OR planning, including hospitalization costs, preferences of surgeons and patients, fitness of operating rooms and additional medical considerations. The determination of the cost matrix $[a_{its}]$ is a difficult task as it involves multiple factors. The design of a simple yet easy to use automatic cost estimation system is an important direction for future research for application of the approach proposed in this paper.

2.3. Emergency surgery

Emergency patients arrive randomly and due to their critical condition they must be operated on as soon as possible on the day of their arrival. In other words, all emergency surgery demand must be met on the day it arises whatever the available capacity.

Emergency patients are clustered into several groups according to their medical requirements. A suitable OR is associated with each group. When one group's emergency patient arrive, (s)he will be operated on in the related OR as soon as the latter becomes available.

In this work, we consider that a random portion of each OR-day's capacity is used to serve emergency patients. Let W_{ts} be the capacity, i.e. total OR time, needed for emergency cases operated in OR-day (s, t). It is assumed to be a random variable with the density function $f_{W_{ts}}(x)$. This distribution can be easily estimated from historical data on emergency capacity requirements for related OR-days in the past.

- *Assumption 2*: We assume that the total OR-time of an OR-day used for emergency patients is a random variable W_{ts} with a given distribution.

2.4. Operating room utilization cost

For each OR-day (s, t) there is a regular capacity T_{ts} , i.e., the number of regular working hours, and an overtime capacity V_{ts} . The sum of these two capacities represents the OR-day's available capacity ($T_{ts} + V_{ts}$). If the aggregated duration of emergency and planned elective surgeries exceeds the regular capacity, then overtime is needed. Let c_{ts} be the cost per time unit of overtime in “OR-day” (s, t). Moreover, if this aggregated duration exceeds the OR-day's available capacity, then additional penalties have to be paid. Let c_{ts} be the additional penalty per time unit for exceeding the available capacity in OR-day (s, t). It is further assumed that the aggregated duration of planned elective surgeries should not exceed the OR-day's available capacity.

- *Assumption 3*: The total duration of planned elective surgeries must not exceed the OR-day's available capacity, i.e., regular capacity plus overtime capacity.

If the total duration of elective surgeries assigned to a given OR-day (s, t) is equal to K, $K \leqslant T_{ts} + V_{ts}$, then the expected overutilization cost related to this OR-day is

$$c_{ts}\, E_{W_{ts}}[(W_{ts} + K - T_{ts})^{+}] + \bar{c}_{ts}\, E_{W_{ts}}[(W_{ts} + K - T_{ts} - V_{ts})^{+}], \qquad (1)$$

where the expectation is with respect to the distributions of W_{ts} , and $(x)^{+} = \max\{0, x\}$.

The major components of an OR's costs are fixed costs, salaries of staff and fixed cost of facilities and equipments (Denton and Gupta, 2003 Denton, B. and Gupta, D. 2003. A sequential bounding approach for optimal appointment scheduling. *IIE Transactions*, 35: 1003–1016. [Taylor & Francis Online], [Web of Science ®], [Google Scholar]), and therefore efficient planning of surgeries requires low OR idle times. In this work, OR underutilization is also considered. Let u_{ts} be the cost per unit of underutilization of OR-day (s, t). The expected underutilization cost can be expressed by

$$u_{ts}\, E_{W_{ts}}[(W_{ts} + K - T_{ts})^{-}], \qquad (2)$$

where $(x)^{-} = \max\{0, -x\}$.

Hence, if the total duration of elective surgeries assigned to OR-day (s, t) is equal to K, then the expected utilization cost $g_{ts}(K)$ of OR s on day t is the sum of overutilization (1) and underutilization (2) costs, that is

$$g_{ts}(K) = c_{ts}E[(W_{ts} + K - T_{ts})^{+}] + \bar{c}_{ts}E[(W_{ts} + K - T_{ts} - V_{ts})^{+}] + u_{ts}E[(W_{ts} + K - T_{ts})^{-}]. \qquad (3)$$

- *Remark 4*: Notice that in the above model, larger than average emergency requirements lead to higher overtime costs and do not lead to surgery cancellation. In practice, significant requirements for emergency surgery or longer than predicted operating times could lead to surgery cancellation. Taking into account surgery cancellation in OR planning is an interesting research direction.

2.5. Mathematical model

The elective surgery planning problem consists in assigning elective patients to OR-days with the best trade-off between elective-patient-related costs, OR overutilization costs and underutilization costs.

Let $X_{its} \in \{0, 1\}$ be the decision variable, with $X_{its} = 1$ if elective case i is performed in OR-day (s, t) and zero otherwise; with the obvious convention that $X_{i,H+1,s} = 1$ implies that elective case i is rejected from the current planning horizon.

The surgery planning problem can now be formulated by the following mathematical model, called the general problem (GP):

$$\text{(GP)} : J^* = \min J(X) = \sum_{i=1}^{N} \sum_{t=e_i}^{H+1} \sum_{s=1}^{S} a_{its} X_{its} + \sum_{t=1}^{H} \sum_{s=1}^{S} [c_{ts} O_{ts} + \bar{c}_{ts} \bar{O}_{ts} + u_{ts} U_{ts}], \quad (4)$$

subject to:

$$O_{ts} = E_{W_{ts}}\left[\left(W_{ts} + \sum_{i=1}^{N} X_{its} d_i - T_{ts}\right)^+\right] \quad \forall t = 1, \ldots, H, \ \ \forall s = 1, \ldots, S, \quad (5)$$

$$\bar{O}_{ts} = E_{W_{ts}}\left[\left(W_{ts} + \sum_{i=1}^{N} X_{its} d_i - T_{ts} - V_{ts}\right)^+\right] \quad \forall t = 1, \ldots, H, \ \ \forall s = 1, \ldots, S, \quad (6)$$

$$U_{ts} = E_{W_{ts}}\left[\left(W_{ts} + \sum_{i=1}^{N} X_{its} d_i - T_{ts}\right)^-\right] \quad \forall t = 1, \ldots, H, \ \ \forall s = 1, \ldots, S, \quad (7)$$

$$\sum_{i=1}^{N} X_{its} d_i \leq T_{ts} + V_{ts}, \forall t = 1, \ldots, H, \quad \forall s = 1, \ldots, S, \quad (8)$$

$$\sum_{t=e_i}^{H+1} \sum_{s=1}^{S} X_{its} = 1 \forall i = 1, \ldots, N, \quad (9)$$

$$X_{its} \in \{0, 1\} \forall i = 1, \ldots, N, \ \ \forall t = 1, \ldots, H, H+1, \quad \forall s = 1, \ldots, S. \quad (10)$$

The expectations are with respect to the distributions of W_{ts} , $(x)^+ = \max\{0, x\}$, and $(x)^- = \max\{0, -x\}$.

The objective function (4) seeks to minimize the sum of the expected OR utilization costs and the elective-patients-related costs. Constraints (5), (6) and (7) determine respectively the expected overtime O_{ts} , available capacity's violation $\bar{O}_{ts}$ and idle time u_{ts} for each OR-day. Constraint (8) ensures that the total duration of planned elective surgeries does not exceed the OR-day's available capacity. Constraint (9) guarantees that each elective case is assigned exactly to one OR-day. Constraint (10) is the integrity constraint.

The elective surgery planning problem (GP) is a stochastic combinatorial problem. It is NP-hard which can be proved by using a polynomial transformation of the three-partition problem. In the next section, we present a column generation approach to find a near-optimal solution for (GP).

3. Solution methodology

The main steps of the column generation approach proposed in this section are the following ones: (i) derive a column-oriented reformulation of the planning problem with a very large number of columns (variables), called the master problem; (ii) solve the Linear Programming (LP) relaxation of the master problem using column generation; (iii) derive a feasible plan from the solution of the LP-relaxation of the master problem; (iv) improve the feasible plan using local optimization.

3.1. A column generation approach

Let us first present a column formulation for the planning problem. A column represents a plan for one OR in one day. In this formulation, a “plan” includes an OR- day and a set of patients assigned to it. A plan p is defined by the following binary variables:

$$y_{ip} = \begin{cases} 1 & \text{if patient } i \text{ is assigned to plan } p, \\ 0 & \text{otherwise.} \end{cases}$$

$$z_{tsp} = \begin{cases} 1 & \text{if plan } p \text{ is assigned to OR-day } (s,t), \\ 0 & \text{otherwise.} \end{cases}$$

Then the plan p can be represented by $[y_p , z_p]=[(y_{1p} , \ldots, y_{Np}), (z_{1p} , \ldots, z_{(H\times S)p})]$. The first N entries, given by y_p , represent the set of patients assigned to the plan. The next $H\times S$ entries, i.e., z_p , indicate to which OR-day(s) the plan is assigned.

Let Ω be the set of all possible feasible plans which is huge but finite. A plan p is feasible if: (i) it is assigned to exactly one OR-day (i.e.,$\sum_{t,s} z_{tsp} =1$); (ii) it respects patients' earliest periods ($y_{ip} \times z_{tsp} =0 \sim \forall t<e_i$); and (iii) the OR-day capacity is satisfied, i.e., $(\sum_i y_{ip}d_i)z_{tsp} \leqslant T_{ts} +V_{ts}$, $\forall t, s$.

Note that the set of plans defined above concerns only OR-days over the planning horizon and no plan in Ω is associated to the virtual period H+1.

The expected cost C_p of plan p can be expressed as follows:

$$C_p = \sum_{t,s} z_{tsp} \left[\sum_i y_{ip}\, a_{its} + g_{ts} \left(\sum_{i=1}^{N} y_{ip}\, d_i \right) \right], \qquad (11)$$

where $g_{ts} (\sum_{i=1}^{N} y_{ip}d_i)$ is the expected utilization cost of OR s on day t defined in Equation (3), with an aggregated duration of elective surgeries equal to $\sum_{i=1}^{N} y_{ip}d_i$. The plan's cost is composed of two parts: elective-patient-related costs and OR-related costs (overtime cost, underutilization costs and available capacity violation penalty).

The planning problem can now be seen as a problem of selecting a set of feasible plans. Let λ_p for $p \in \Omega$ be a binary decision variable indicating whether a feasible plan p is selected ($\lambda_p = 1$) or not ($\lambda_p = 0$). The planning problem can now be formulated as follows:

$$\min \sum_{p\in\Omega} C_p\lambda_p + \sum_i a_{iH+1}\left(1 - \sum_{p\in\Omega} y_{ip}\lambda_p\right) + \sum_{t,s} g_{ts}(0) \times \left(1 - \sum_{p\in\Omega} z_{tsp}\lambda_p\right), \quad (12)$$

subject to

$$\sum_{p\in\Omega} y_{ip}\lambda_p \le 1 \quad \forall i = 1,\ldots,N, \quad (13)$$

$$\sum_{p\in\Omega} z_{tsp}\lambda_p \le 1 \quad \forall t = 1,\ldots,H, \forall s = 1,\ldots,S, \quad (14)$$

$$\lambda_p \in \{0,1\}, \quad \forall p \in \Omega. \quad (15)$$

Note that patients assigned to selected plan p with $\lambda_p = 1$ are assigned to OR-days over the planning horizon and the remaining patients are assigned to the virtual period H+1, i.e., they are excluded from the current planning horizon. Furthermore, some OR-days may not be assigned to any selected plan, i.e., no elective patients.

The objective function is again the minimization of costs, but now expressed in terms of the new λ_p variables. It is composed of three parts. The first part corresponds to the cost of selected plans. The second part is the cost incurred by patients assigned to the virtual period H+1, i.e., excluded from the current planning and not belonging to any selected plan. Finally, the third part represents costs related to the OR-days that are not associated with any selected plan and hence do not receive any elective patients. In other words, this OR-day is for emergency cases only. Constraints (14) guarantee that each patient is planned in at most one OR-day within the planning horizon. Constraints (14) ensure that at most one feasible plan is selected for each OR-day.

Let *[Ctilde]* $_p = C_p - \sum_i a_{iH} + 1\, y_{ip} - \sum_{t,s} g_{ts}(0)\, z_{tsp}$. By regrouping relevant terms, the column-oriented formulation of the planning problem, called the master problem (MP), can be restated as follows:

$$(\text{MP}) : \min \sum_i a_{iH+1} + \sum_{t,s} g_{ts}(0) + \sum_{p\in\Omega} \tilde{C}_p\lambda_p, \quad (16)$$

subject to Equations (13), (14) and (15).

Note that the master problem is an integer LP problem, whereas the initial formulation (GP) has a non-linear objective function. The non-linear terms of (GP) are incorporated in the cost coefficient columns.

The master problem contains a huge number of columns and cannot be solved directly. The linear relaxation of (MP) (called the linear master problem (LMP)) with the integrity constraint (15) being replaced by $\lambda_p \ge 0$, $\forall p\in\Omega$ is considered and solved by column generation. Since the number of columns is very large, a restricted master problem (RMP) that considers only a subset $\Omega' \subset \Omega$ of the columns is solved, and additional columns are generated only as needed as described in the following.

Let π_i and $\pi_{t,s}$ be the optimal dual variables to (RMP), associated respectively with constraints (13) and (14). The pricing problem:

$$\sigma_p = \min_{\{y_{ip}, z_{tsp}\}} \tilde{C}_p - \sum_i \pi_i y_{ip} - \sum_{t,s} \pi_{ts} z_{tsp}$$
$$\text{subject to } p \in \Omega,$$

identifies a column with minimum reduced cost. If the minimum reduced cost is less than zero, the corresponding column, i.e., plan $[y_p, z_p]$, can be added to (RMP) for a new iteration; if the minimum reduced cost is greater or equal to zero, then the current optimal solution for (RMP) is also optimal for (LMP).

Clearly the optimal cost of (LMP) is a lower bound of the optimal cost of the master problem (MP) and hence a lower bound of the optimal cost of the elective patient planning problem (GP).

3.2. The pricing problem

The pricing problem consists in finding the plan *p* (column) with the minimum reduced cost. From the definition of feasible plans, the pricing problem can be formulated as follows:

$$\min \tilde{C}_p - \sum_{i=1}^{N} \pi_i y_{ip} - \sum_{t=1}^{H} \sum_{s=1}^{S} \pi_{ts} z_{tsp}, \qquad (17)$$

subject to

$$y_{ip} \times z_{tsp} = 0 \quad \forall t < e_i, \qquad (18)$$

$$\left(\sum_i y_{ip} d_i \right) z_{tsp} \le T_{ts} + V_{ts} \quad \forall t, s, \qquad (19)$$

$$\sum_{t,s} z_{tsp} = 1, \qquad (20)$$

$$y_{ip} \in \{0, 1\}, z_{tsp} \in \{0, 1\} \quad \forall i, t, s, \qquad (21)$$

where

$$\tilde{C}_p = \sum_{t,s} z_{tsp} \left[\sum_i y_{ip} a_{its} + g_{ts} \left(\sum_{i=1}^{N} y_{ip} d_i \right) \right]$$
$$- \sum_i a_{iH+1} y_{ip} - \sum_{t,s} g_{ts}(0) z_{tsp}.$$

The objective function (17) gives the reduced cost with respect to the dual solution of the current (RMP), given by π_i and π_{ts}. Constraints (18) ensure that no patient is planned before his/her earliest period. Constraints (19) guarantee that the total duration of elective surgeries assigned to the plan do not exceed the OR-day's available capacity (regular capacity plus overtime capacity). Constraints (20) ensure that the plan is assigned to exactly one OR-day.

The pricing problem can be decomposed into $H \times S$ subproblems, one for each OR-day. The pricing can be solved by solving the following column generation subproblem ($CG_{t,s}$) for each OR-day:

$$(\mathrm{CG}_{t,s}) : \min \sum_{i=1}^{N} (a_{its} - a_{iH+1} - \pi_i) y_{ip}$$

$$+ g_{ts}\left(\sum_{i=1}^{N} y_{ip} d_i\right) - g_{ts}(0) - \pi_{ts}, \qquad (22)$$

subject to

$$y_{ip} = 0 \ \ \forall i \text{ with } t < e_i, \qquad (23)$$

$$\sum_i y_{ip} d_i \le (T_{ts} + V_{ts}) \ \ \forall t, s, \qquad (24)$$

$$y_{ip} \in \{0, 1\} \ \ \forall i = 1, \ldots, N. \qquad (25)$$

The pricing problem defined above finds the column with the minimum reduced cost. Therefore, if a column with a negative reduced cost exists it will always be identified by the pricing problem. This guarantees that the optimal solution to the linear program (LMP) will be found (Barnhart *et al.*, 1998 Barnhart, C., Johnson, E. L., Nemhauser, G. L., Savelsbergh, M. W. P. and Vance, P. H. 1998. Branch-and-price: column generation for solving huge integer programs. *Operations Research*, 46: 316–329. [CrossRef], [Web of Science ®], [Google Scholar]). However, column generation works with any column with negative reduced cost.

Three column generation strategies will be used in this paper. The first strategy corresponds to the classical Dantzig rule (Lübbecke and Desrosiers, 2005 Lübbecke, M. E. and Desrosiers, J. 2005. Selected topics in column generation. *Operations Research*, 53: 1007–1023. [CrossRef], [Web of Science ®], [Google Scholar]), and it is to select among all columns the one with the most negative reduced cost (best-negative strategy). The second strategy is to select for each OR-day the most negative reduced cost column (all-negative strategy). The third strategy consists in selecting the first encountered column with a negative reduced cost; at each iteration OR-days' subproblems are solved in a random order. Obviously, for all three strategies, columns with a non-negative reduced cost are never added to the (RMP).

The remainder of this section is devoted to the solution of the column generation subproblem ($\mathrm{CG}_{t,s}$) defined in Equations (22) to (25). The subproblem ($\mathrm{CG}_{t,s}$) consists in determining the set of patients to be operated on in the OR-day (*s*, *t*), for a given set of π_i. A dynamic programming algorithm is proposed here to solve the subproblem.

In Equation (22) the term $-g_{ts}(0) - \pi_{ts}$ is constant, and hence can be dropped from the criteria. For each period *t*, let $I_t = \{i / e_i \leqslant t\}$ be the set of patients that can be operated on in period *t*. By setting $y_{ip} = 0 \sim \forall i \notin I_t$, the column generation subproblem can be reformulated as follows:

$$\min \sum_{i \in I_t} \tilde{a}_{its} y_{ip} + g_{ts}(K), \qquad (26)$$

subject to

$$\sum_{i \in I_t} y_{ip} d_i = K, \qquad (27)$$

$$0 \le K \le K_{\max}, \qquad (28)$$

$$y_{ip} \in \{0, 1\} \ \ \forall i \in I_t, \qquad (29)$$

where $\tilde{a}_{its} = a_{its} - a_{iH} + 1 - \pi_i$ and $K_{max} = T_{ts} + V_{ts}$.

- *Remark 5*: Additional constraints related to surgeon, patient or OR availability can be easily taken into account. For each OR-day (s, t) a set I_{st} of assignable patients can be defined with respect to these additional constraints. It is enough to replace I_t by I_{st} .

For $0 \leq K \leq K_{max}$, let us define the following 0–1 knapsack problem:

$$h(K) = \min_{\{y_{ip}\}} \left\{ \sum_{i \in I_t} \tilde{a}_{its} y_{ip} \text{ subject to } \sum_{i \in I_t} y_{ip} d_i = K \right\}. \quad (30)$$

Then the subproblem ($CG_{t,s}$) is equivalent to

$$\min_{0 \leq K \leq K_{\max}} \{h(K) + g_{ts}(K)\}. \quad (31)$$

g_{ts} (K) can be evaluated (computed) by numerical integration over the distribution of W_{ts} . The different values of $h(K)$ can be determined as follows by using dynamic programming to solve the 0–1 knapsack problem (30) with $K = K_{max}$.

Given a pair of integers m ($0 \leq m \leq |I_t|$) and K ($0 \leq K \leqslant K_{max}$), let

$$h_m(K) = \min \left\{ \sum_{i=1}^{m} \tilde{a}_{its} y_{ip} \text{ subject to } \sum_{i=1}^{m} y_{ip} d_i = K \right\}. \quad (32)$$

Dynamic programming consists in considering $|I_t|$ stages (for m increasing from one to $|I_t|$) and computing at each stage the values h_m (K) (for K increasing from zero to K_{max}) using the following recursion:

$$h_m(K) = \begin{cases} h_{m-1}(K) & \text{for } 0 \leq K \leq d_m - 1 \\ \min(h_{m-1}(K), & \text{for } d_m \leq K \leq K_{\max} \\ \quad h_{m-1}(K - d_m) + \tilde{a}_{mts}) & \end{cases}$$

by starting with

$$h_1(K) = \begin{cases} 0 & \text{for } K = 0, \\ \tilde{a}_{1ts} & \text{for } K = d_1, \\ +\infty & \text{for } K = 1 \ldots K_{\max}, K \neq d_1 \end{cases}$$

The optimal solution to subproblem ($CG_{t,s}$) is the solution corresponding to $h_{|I_t|}$ (K^*), with

$$K* = \arg\min_{0 \leq K \leq K_{\max}} h_{|I_t|}(K) + g_{ts}(K). \quad (33)$$

The solution method used for solving column generation subproblems is similar to the dynamic programming method for knapsack problems except that the additional non-linear cost is taken into account at the last stage.

3.3. Constructing a near-optimal solution

Given λ_p the optimal solution of the linear master problem, a solution X_{its} of the general problem (GP) can be derived by the following expression:

$$X_{its} = \sum_{p \in \Omega} y_{ip} z_{tsp} \lambda_p. \quad (34)$$

If the λ_p variables are also all integers, then the X_{its} variables are also integers. That is to say, any integer solution to (LMP) has a corresponding feasible solution to (GP). However, if the solution to (LMP) is fractional (i.e., not feasible), then the corresponding solution to (GP) is also fractional.

Therefore, unless the optimal solution to (LMP) is integral, the solution provided by column generation is not feasible for (MP), and the corresponding solution for (GP) is also infeasible. Heuristic methods will be developed to construct a “good” feasible solution based on the optimal solution to (LMP). We first construct a feasible solution and then attempt to improve it using one or several local optimization methods.

3.3.1. Deriving a feasible solution

The straightforward method to derive a feasible solution to (GP) is to solve (MP) when restricted to columns generated when solving (LMP). The following presents two other methods.

3.3.1.1. *Complete reassignment method*. Given a fractional optimal solution to (LMP) $\{\lambda_p\}$, the main steps of the method are as follows. First, patients assigned to selected plans with $\lambda_p = 1$ are assigned to OR-days corresponding to these plans. Let X represent the partial solution formed by these assignments. Then, all other patients will be reassigned one by one and by taking into account patients already planned.

An elective case i which has yet to be assigned is first selected. New solutions $X^{(s,t)}$ are then derived from X by adding the new case i in OR-day (s, t) while keeping the previous assignments unchanged; let $J(X^{(s,t)})$ be the costs of these new solutions. The case i is finally assigned to the OR-day (s_i, t_i) that minimizes $J(X^{(s,t)})$, i.e., $(s_i, t_i) = \operatorname{argmin}_{\{(s,t)\}} J(X^{(s,t)})$. Then, X is replaced by $X^{(s_i, t_i)}$ and the algorithm continues until all elective cases are assigned.

3.3.1.2. *Progressive reassignment method* Let X_{its} be the solution to (GP) that corresponds to the optimal solution to (LMP) i.e., $X_{its} = \sum_{p \in \Omega} y_{ip} z_{tsp} \lambda_p$. An elective case i such that its assignment matrix $[X_{its}]_{ts}$ contains fractional numbers is first selected. We then consider solutions $X^{(s,t)}$ such that case i is reassigned to OR-day (s, t) only while all other assignment variables X_{its}, fractional or not, are kept unchanged. Let $J(X^{(s,t)})$ be the costs of solutions $X^{(s,t)}$.

The case i is finally assigned to the OR-day (s_i, t_i) that minimizes $J(X^{(s,t)})$, i.e., $(s_i, t_i) = \operatorname{argmin}_{\{(s,t)\}} J(X^{(s,t)})$. Obviously, $t_i \geq e_i$ and $\sum_{j=1}^{N} X_{jt_i s_i} d_i \leq T_{t_i s_i} + V_{t_i s_i}$. The solution X is then replaced by $X^{(s_i, t_i)}$, and the process continues for other fractional assignments.

3.3.2. Improvement of a feasible solution

Starting from a feasible solution, the following optimization methods will be used to improve the solution.

3.3.2.1. *Local optimization of elective cases*. The solution is iteratively improved by reassigning elective patients. At each iteration, we compute for each patient i the largest improvement that can be achieved by reassigning it to another OR-day (or to period $H + 1$), while keeping all others patients' assignments unchanged. Clearly, the reassignment must satisfy the patient's earliest period and the OR-day's available capacity. Then we choose among all patients the one that leads to the largest improvement, and we reassign that patient. The iterations continue until the solution cannot be improved any more.

3.3.2.2. *Pair-wise exchange of elective cases*. The principle of this method is to iteratively improve the solution by exchanging the assignment of a couple of elective patients.

At each iteration, patients are considered one by one. At each step, we consider a patient i, we determine the patient j that leads to the largest improvement if the assignments of patients i and j are exchanged, without modifying the assignment of all other patients, and while satisfying constraints related to patients' earliest periods and OR-days' available capacities. Then, we exchange the assignments of patients i and j. The next step considers another patient and so on.

At the next iteration, patients are again considered one by one. The process stops when the solution cannot be improved any more.

3.3.2.3. *Period-based re-optimization*. We consider all the patients assigned to a given OR-day (s, t) and all rejected patients (i.e., assigned to period $H + 1$) and reoptimize the partition of these patients between the OR-day (s, t) and period $H + 1$.

Let I_{st} be the set of elective patients assigned to the OR-day (s, t), and I_{H+1} be the set of elective patients assigned to the period $H + 1$, i.e., excluded from the current planning. Let $Z_i \in \{0, 1\}$ for $i \in I_{ts} \cup I_{H+1}$ be the decision variable indicating whether a patient is assigned to the OR-day (s, t) $(Z_i = 1)$ or to the period $H + 1$ $(Z_i = 0)$ in the modified solution. The reoptimization problem related to the OR-day (s, t) can now be formulated as follows:

$$(RP_{st}):\ \min \sum_{i \in I_{st} \cup I_{H+1}} a_{its} Z_i + \sum_{i \in I_{st} \cup I_{H+1}} a_{iH+1}(1 - Z_i) + g_{ts}\left(\sum_{i \in I_{st} \cup I_{H+1}} d_i Z_i\right),$$

subject to

$$\sum_{i \in I_{st} \cup I_{H+1}} d_i Z_i \leq T_{ts} + V_{ts}, \qquad (35)$$

$$Z_i \in \{0, 1\}\ \ \forall i \in I_{ts} \cup I_{H+1}. \qquad (36)$$

The objective function is composed of the costs related to: (i) elective patients assigned to OR-day (s, t); (ii) excluded elective patients; and (iii) the utilization of OR s in period t. Constraint (35) is the available capacity constraint.

By a simple arrangement of terms the above problem can be restated as follows:

$$\min \sum_{i \in I_{st} \cup I_{H+1}} a_{iH+1} + \sum_{i \in I_{st} \cup I_{H+1}} (a_{its} - a_{iH+1}) Z_i + g_{ts}\left(\sum_{i \in I_{st} \cup I_{H+1}} d_i Z_i\right),$$

subject to Equation (35) and (36).

This is similar to the column generation subproblem (CG_{ts}) and can be solved by using the dynamic programming method presented in Section 3.2.

At each iteration, an OR-day (*s*, *t*) is selected and the related reoptimization problem (RP_{st}) is solved, and elective patients are reassigned according to the solution of (RP_{st}). At the next iteration another OR-day is considered and the process continues until no improvement is possible. OR-days are considered in their chronological order.

3.4. Combination of heuristic methods

The solution methodology proposed in this paper is composed of three stages. In the first stage, the linear master problem (LMP) is solved by using column generation with the restricted master problem (RMP) solved via CPLEX LP and the column generation subproblems via dynamic programming. In the second stage a feasible solution to the general problem (GP) is derived from the optimal solution to (LMP). This task can be accomplished by using one of the following three methods: (i) integer programming (CPLEX IP) to find the best feasible solution based on the column pool obtained after solving LMP; (ii) Complete Reassignment (CR); and (iii) Progressive Reassignment (PR). The third stage aims at improving the feasible solution by using a combination of the following heuristics: (LO) Local Optimization of elective cases, (PE) Pair-Wise exchange of elective cases, and (PB) Period-Based reoptimization.

In the next section, seven combinations of these methods will be tested. These combinations are summarized in Table 1, where each row represents a combination denoted by M#.

Table 1 The different combinations

Method	**Solving LMP**	**Deriving a feasible solution**	**Improving the solution**
M1	Column	CPLEX IP	
M2	generation:	CR	LO
M3		PR	LO
M4	CPLEX LP	PR	PE, PB
M5	+	PR	PE, LO, PB
M6	Dynamic	PR	LO, PB, PE
M7	Programming	PR	LO, PE

In order to evaluate the performance of the different methods, we use the relative optimality gap (GAP) defined as follow:

$$GAP = (UB - LB)/UB,$$

where *UB* is the cost of the feasible solution, which represents an upper bound on the optimal cost *J** of (GP), and *LB* is the optimal cost of (LMP) provided by column generation which is a lower bound of *J**.

When solving (LMP), three columns generation schemes, best-negative, all-negative, and first-negative strategies (Section 3.2), can be used to determine new columns to add at each iteration.

4. Computational results

The numerical experiments presented in this section were carried out on a 2.8 GHz Pentium 4 PC with the memory of 512 Mb, and running Windows XP. The algorithms were implemented in MS Visual C++ and linked with the CPLEX 8.1 optimization library.

4.1. Data

Two classes of problems were considered. For the first class, elective patients can be assigned to any OR without any additional cost, i.e., for each elective case i the costs a_{its} depend only on the date of surgery. For the second class, the ORs are of different types and are allocated to different specialties. We suppose that there are three specialties, and the ORs are equally allocated to them. For example, if the operating theater is composed of six ORs, then two ORs are allocated to each specialty. ORs allocated to the same specialty are considered to be identical. However, ORs allocated to different specialties are considered to be different due to their location or special medical equipment. For this class of problems, one specialty's elective surgery can be performed in an OR allocated to another specialty, but at an additional cost. Since each specialty may require special medical equipment or preliminary preparations in the OR these additional costs are due to an extra amount of set up or switching time.

Problems with three, six, nine and 12 ORs will be considered for the two problem classes. Problem instances are generated randomly according to the following scheme.

The number of periods H is set equal to five (one week). The OR-days' regular capacities T_{ts} and overtime capacities V_{ts} are respectively equal to 8 and 3 hours. The overtime cost c_{ts} , the penalty for exceeding the OR-day available capacity $\bar{c}_{ts}$, and the underutilization cost u_{ts} are respectively 500 €/hour, 3000 €/hour and c_{ts} /1.75. The random capacity used in each OR-day by emergency surgery W_{ts} is assumed to be exponentially distributed with a mean of 2 hours ($E[W_{ts}]$ = 2 hours).

The durations of elective surgeries d_i are randomly and uniformly generated from the interval [0.5 hour, 3 hours]. These durations are multiples of 5 minutes.

For each elective case i we randomly select its release period e_i from the set $\{1,\ldots,H\}$. To take into account cases with $e_i = 1$ that were postponed from the previous planning horizon, we introduce a new variable e_i' the *effective* earliest period of i (or *effective* release period). e_i' can take negative values. Earliest dates e_i are generated in two steps as follows. First, we generate for each case i the *effective* earliest period e_i'. The e_i' values are integer numbers randomly and uniformly generated from the set $\{-2, \ldots, H\}$. Then, cases with zero or negative e_i' will have e_i equal to one, while the others will have e_i equal to e_i' ($e_i = 1$ if $e_i' < 1$; $e_i = e_i'$ otherwise).

The elective-patient-related costs are generated in two different ways according to which problem class the instance belongs. For the first class, a_{its} the cost of assigning patient i to OR-day (s, t) is independent of the OR *s* and depends only on the period *t*. Thus, for each patient *i* and period *t*, we set $a_{its} = a_{its}$ ' for any *s*, $s \in \{1, \ldots, S\}$.

The patient-related costs a_{its} are assumed to be increasing in *t* for each *i*, and defined as follows:

and

The constant *c* can be interpreted as a penalty per day of waiting to be scheduled, or as the hospitalization cost per day if patient *i* was hospitalized in period e_i '. The quantity 2 × c within $a_{i(H+1)}$ represents the penalty of excluding the case from the current planning. The value of *c* is set equal to € 350.

For the second class of problems, patients-related costs are generated as follows. For each elective patient *i*, we have:

if OR *s* is allocated to the patient's specialty, and

otherwise. $a_{i(H+1)}$ is as above. The constant *R* represents the additional penalty to be paid if the elective patient is assigned to an unsuitable OR. In the numerical experiments, several values of *R* (€100, €200, €300 and €400) are used for the second class of problems. If it is not precisely mentioned *R* equals €100.

- *Remark 6*: If $R = 0$ then patients can be assigned to any OR with no additional cost. Hence, a problem with identical ORs is equivalent to a problem with non-identical ORs but with *R* set to zero.

The cost parameters c_{ts} (overtime cost), u_{ts} (underutilization cost) and *c* (hospitalization cost) are similar to those used in Jebali *et al.* (2006) Jebali, A., Hadjalouane, A. B. and Ladet, P. 2006. Operating rooms scheduling. *International Journal of Production Economics*, 99: 52–62. [CrossRef], [Web of Science ®], [Google Scholar] and Guinet and Chaabane (2003) Guinet, A. and Chaabane, S. 2003. Operating theatre planning. *International Journal of Production Economics*, 85: 69–81. [CrossRef], [Web of Science ®], [Google Scholar], which were observed in French hospitals. A large value is chosen for the penalty $\bar{c}_{ts}$ in order to penalize exceeding the OR-day's available capacity. The penalty *R* is introduced essentially for experimentation purposes. Using this parameter, we can modify patients' assignment costs and perform some sensitivity analysis.

Elective cases were generated one by one until the sum of the operating time for all elective cases exceeds a given τ% of the regular capacity of all ORs in the entire planning horizon. In other words, the number of elective cases is determined such that the workload of ORs due to elective cases is τ % of the regular capacity of the entire planning horizon. In these experiments we consider problems with τ equal to 75 and 100%.

With $E[W_{ts}]$ equal to 2 hours and T_{ts} equal to 8 hours, the workload of ORs due to emergency cases is 25% of the regular capacity of the entire planning horizon. So when τ = 75%, elective cases and emergency surgeries sum up to an average of 100% of the regular capacity of the ORs, and when τ = 100% elective and emergency surgeries sum up to 125% of the ORs' regular capacity. Hence, with τ equal to 75 or 100% we have a high workload for the ORs which makes the surgery planning problem more difficult.

For the two problem classes, the problem sizes are determined by the number of ORs and the τ percentages. For the experiments we used the following sizes (S = 3, 6, 9, 12 and τ = 75, 100%) for the two problem classes.

4.2. Evaluation and comparison of different methods

We first compare the performance of the seven methods (combinations) presented in Table 1. These methods are tested with various problem sizes (S= 3, 6, 9, 12 and τ = 75%) and for the two problem classes (identical ORs and non-identical ORs with R = 100). For the pricing problem, the classical scheme, i.e., the best-negative strategy, is used. The (RMP) starts with an initial subset Ω of columns generated randomly, such that the corresponding solution of (GP) is a feasible one.

All computational results presented are the average of ten randomly generated instances for the size and class under consideration. The results can be found in Table 2 and Table 3. The results include, for each problem size, the average number of elective patients (number of patients), and the performances of the seven methods (combinations). For each method, we provide the relative optimality gap (GAP), the CPU time needed to solve (LMP) (CPU Linear), the CPU time needed to construct a near-optimal solution (CPU Integer) and the total CPU time (CPU).

Table 2 Results for identical ORs with τ = 75%

Number of ORs	**Number of patients**	**Method**	**GAP (%)**	**CPU Linear (sec)**	**CPU Integer (sec)**	**CPU (sec)**
3	53.9	M1	18.31	8.60	43.00	51.60
		M2	3.78		0.15	8.76
		M3	2.72		0.12	8.72
		M4	2.83		0.05	8.65
		M5	2.13		0.08	8.68
		M6	1.52		0.17	8.77
		M7	1.54		0.14	8.75
6	106.9	M1	24.80* *Based on two instances. View all notes	42.76	7820.29	7863.05
		M2	2.81		0.93	43.69
		M3	2.27		0.72	43.48
		M4	1.74		0.15	42.90
		M5	1.69		0.25	43.00
		M6	1.40		1.08	43.84
		M7	1.40		0.80	43.56
9	160.1	M1	—	138.32	>8000	>8000
		M2	3.25		2.58	140.90
		M3	2.26		2.15	140.47
		M4	1.51		0.31	138.63
		M5	1.44		0.61	138.93
		M6	1.15		2.41	140.73

Number of ORs	Number of patients	Method	GAP (%)	CPU Linear (sec)	CPU Integer (sec)	CPU (sec)
		M7	1.15		2.34	140.66
12	211.8	M1	—	344.30	>8000	>8000
		M2	2.86		6.29	350.60
		M3	1.99		4.70	349.00
		M4	1.33		0.57	344.87
		M5	1.29		1.13	345.44
		M6	0.88		5.45	349.75
		M7	0.88		5.05	349.35

*Based on two instances.

Table 3 Results for non-identical ORs with τ = 75%

Number of ORs	Number of patients	Method	GAP (%)	CPU Linear (sec)	CPU Integer (sec)	CPU (sec)
3	53.9	M1	0.95	17.34	1.06	18.40
		M2	3.91		0.06	17.40
		M3	1.56		0.06	17.40
		M4	2.29		0.04	17.38
		M5	1.30		0.08	17.43
		M6	0.87		0.12	17.46
		M7	0.93		0.09	17.43
6	106.9	M1	—	76.99	>8000	>8000
		M2	5.68		0.89	77.88
		M3	2.47		0.54	77.53
		M4	2.05		0.15	77.14
		M5	1.86		0.29	77.28
		M6	1.53		0.73	77.72
		M7	1.53		0.63	77.62
9	160.1	M1	—	190.40	>8000	>8000
		M2	5.83		2.31	192.70
		M3	2.82		2.00	192.40
		M4	2.13		0.32	190.72
		M5	2.06		0.62	191.02
		M6	1.58		2.66	193.06
		M7	1.57		2.18	192.58
12	211.8	M1	—	402.84	>8000	>8000
		M2	5.68		5.91	408.76
		M3	3.16		4.68	407.53
		M4	2.17		0.58	403.42
		M5	2.10		1.38	404.22
		M6	1.70		6.01	408.85
		M7	1.70		5.03	407.87

The computational results show that method M1 is very time-consuming. For problems with a size greater than or equal to six ORs, the solution cannot be found within a practical time budget, “CPU Integer” exceeds 2 hours. The other methods construct near-optimal solutions in less than 7 seconds. Moreover, the gap values calculated by M1 are very bad compared with the other methods. This suggests that solving the (MP) based on the column pool obtained after solving (LMP) can be very poor and very time-consuming.

The results in Table 2 and Table 3 also show that method M3 using PR to derive a feasible solution clearly outperforms method M2 using CR. This is due to the fact that PR preserves the structure of the optimal solution to (LMP) whereas CR only takes into account the integral part of the optimal solution to (LMP).

From Table 2 and Table 3 it can also be observed that methods M3 to M7 have comparable performances. However, M6 and M7 are better than the others. Problems with non-identical ORs need a longer computation time to solve (LMP) (CPU Linear) and the GAP is slightly larger when comparing with problems with identical ORs.

In the next experiments, the performances of methods M2 to M7 are compared for problems with size (S = 3, 6, 9, 12 and τ = 100%) for the two problem classes. Computational results are presented in Table 4, and they are based on ten randomly generated instances for each size and each class under consideration. The results include, for each problem size, the average number of elective patients (Number of patients), and the relative optimality gap (GAP) of the six methods. Computation times of these methods are almost the same as those presented in Table 2 and Table 3.

Table 4 The GAP for problems with τ = 100%

Number of ORs	**Number of patients**	**Method**	**Identical ORs GAP (%)**	**Non-identical ORs GAP (%)**
3	71.40	M2	3.12	3.22
		M3	2.85	1.54
		M4	2.67	1.28
		M5	2.63	1.07
		M6	2.04	0.76
		M7	2.02	0.78
6	142.30	M2	2.77	5.52
		M3	2.28	2.46
		M4	2.00	2.30
		M5	1.95	2.15
		M6	1.31	1.53
		M7	1.35	1.57
9	211.80	M2	2.63	4.86
		M3	2.16	3.36
		M4	1.71	2.56
		M5	1.71	2.52
		M6	1.18	1.93
		M7	1.20	1.89
12	283.10	M2	2.45	4.37

Number of ORs	Number of patients	Method	Identical ORs GAP (%)	Non-identical ORs GAP (%)
		M3	2.42	3.38
		M4	1.82	2.39
		M5	1.77	2.35
		M6	1.35	1.82
		M7	1.35	1.84

From Table 4, methods M6 and M7 remain better than the others. However, M6 slightly outperforms M7. This can be explained by the PB reoptimization method incorporated in M6 which is useful when there are many elective patients excluded from the current planning, i.e., assigned to period H+1. These results also show that GAP value is generally higher compared with problems with τ = 75%.

4.3. Comparison of different pricing strategies

In all the above numerical experiments, over 65% of the computation time (CPU Linear) needed to solve (LMP) is spent in solving the pricing problem. In the next experiments, for the set of instances with τ = 75%, we compare the impact of the pricing strategies best-negative, all-negative and first-negative on solving (LMP). A specific heuristic method (combination) for deriving feasible solutions is not considered. Computational results are presented in Table 5, and they are based on ten randomly generated instances for each size and each class under consideration. These instances are the same as those, with τ= 75%, used in the previous experiments. The (RMP) starts with an initial subset Ω of columns generated randomly, such that the corresponding solution of (GP) is a feasible one.

Table 5 Performances of pricing strategies (CPU times in seconds)

	Identical ORs				**Non-identical ORs**			
Number of ORs	**3**	**6**	**9**	**12**	**3**	**6**	**9**	**12**
Best-negative								
Number of columns	284.70	562.78	957.30	1288.80	379.90	752.00	1109.20	1468.80
CPU Pricing	6.09	31.36	101.18	238.23	8.73	46.71	126.40	274.03
CPU RMP	2.52	11.39	37.14	106.07	8.61	30.28	64.00	128.81
CPU Linear	8.60	42.76	138.32	344.30	17.34	76.99	190.40	402.84
All-negative								
Number of columns	1034.10	4039.20	9419.40	16694.40	747.60	2637.00	5863.20	9950.00
CPU Pricing	1.52	7.58	22.62	45.27	1.18	4.92	13.83	27.62
CPU RMP	1.40	7.94	26.24	49.71	1.34	6.25	17.58	37.17
CPU Linear	2.92	15.52	48.86	94.98	2.52	11.17	31.42	64.79
First-negative								
Number of columns	400.20	855.00	1269.20	1655.40	684.60	1240.40	1641.80	2210.40
CPU Pricing	9.08	51.14	144.53	302.08	15.35	72.77	183.55	396.77
CPU RMP	2.79	15.77	46.29	102.71	4.09	21.07	57.32	130.44
CPU Linear	11.87	66.92	190.82	404.79	19.43	93.84	240.88	527.21

The results in Table 5 include, for each strategy, the number of generated columns (number of columns), the CPU time to solve the pricing problem (CPU Pricing), the CPU time to solve the (RMP) (CPU RMP), and the total CPU time to solve the (LMP) (CPU Linear); the time unit is the second. The average number of elective patients for each problem size is presented in Table 2 and Table 3.

The computational results show that the all-negative strategy outperforms the two others strategies and leads to significant reductions in computation time.

Using the all-negative or the first-negative strategies the number of generated columns is larger than using the best-negative strategy. Compared with the best-negative strategy, the all-negative strategy reduces the computation time for solving (LMP) (CPU Linear), but the first-negative strategy increases the computation time.

Solving the pricing problem involves the solution of $H \times S$ column generation subproblems (CG_{ts}) that are computationally intensive. With the best-negative strategy only one column is added to (RMP) at each iteration. Using the all-negative strategy, all columns with a negative reduced cost are added to (RMP) at each iteration. This scheme does not affect the time required to solve the pricing problem, but it increases the time needed to solve (RMP) (per iteration), since a larger (RMP) has to be solved. However, it decreases the number of iterations, which leads to a smaller overall computation time (CPU Linear). The use of the first-negative strategy reduces the computation time required to solve the pricing problem. However, the number of iterations increases which leads to a higher overall computation time.

We also note that for the best-negative and first-negative strategies, problems with non-identical ORs need a larger computation time (CPU Linear) and a larger number of columns compared to those with identical ORs. However, the opposite is true when using the all-negative strategy.

In the next experiments, we evaluate the performances of methods M1 and M6 using the all-negative strategy when solving (LMP) for the case with $\tau = 75\%$. Table 6 provides the computational results. Comparing with results in Table 2 and Table 3, it can be observed that the computation time of method M6 is considerably reduced by using the all-negative strategy in the solution of (LMP). Problems with non-identical ORs need a smaller CPU time compared to those with identical ORs, which is the opposite of the results presented in Table 2 and Table 3. The GAPs are slightly different from those presented in Table 2 and Table 3 where LMP was solved using the best-negative strategy. This can be explained by the non-uniqueness of the optimal solution of (LMP), which leads to different feasible solutions and hence to different GAPs. As far as method M1 is concerned, GAPs are smaller but still large compared to the GAPs of method M6. This is a logical consequence, since more columns are now considered when solving the integer program for deriving a feasible solution, which leads to a better feasible solution. However, the computation time (CPU Time) considerably increases for problems with three identical ORs. For problems with more than three ORs no feasible solutions can be derived by using CPLEX IP to solve (MP) due to the large number of generated columns.

Table 6 Performances of methods M1 and M6 with the all-negative strategy

		Identical ORs		Non-identical ORs	
Method	**Number of ORs**	**GAP (%)**	**CPU Time (sec)**	**GAP (%)**	**CPU Time (sec)**
M1	3	4.95	1141.46	0.43	2.85
	6, 9 and 12	—	—	—	—
M6	3	1.78	3.05	0.77	2.6
	6	1.01	16.34	1.45	11.87
	9	0.81	51.36	1.53	33.58
	12	1.19	99.83	1.57	70.00

To sum up, method M6 provides solutions within 2% of the optimum in a short computation time even for problems with a large size. Although, the results in Table 6 do not include all the proposed methods, we observed that methods M4, M5 and M7 have a similar behavior to method M6.

In the next experiments, we investigate the sensitivity of performance measures (GAP and computation time) of method M6 with respect to patients' assignment costs. For this purpose, problem instances with $\tau = 75\%$, non-identical ORs and various values of R ($R = 200$, 300 and 400) are considered. Numerical results are presented in Table 7. They are based on ten randomly generated instances for each problem size and each value of R. For each problem size under consideration, instances corresponding to each value of R differ only by their patients' assignment costs. Results in Table 7 include the GAP and the computation time (CPU Time). The average number of elective patients corresponding to each problem size is presented in Table 2 and Table 3.

Table 7 Performances of method M6 with different values of R

	***R* = 200**		***R* = 300**		***R* = 400**	
Number of ORs	**GAP (%)**	**CPU Time (sec)**	**GAP (%)**	**CPU Time (sec)**	**GAP (%)**	**CPU Time (sec)**
3	0.63	2.73	0.26	3.46	0.34	2.36
6	1.70	10.75	1.40	10.41	1.44	10.42
9	1.66	29.76	1.31	28.33	1.19	28.50
12	1.30	62.41	1.60	61.10	1.24	55.36

From the results in Table 7, it can be observed that the performance of method M6 is not sensitive to changes that affect the value of patients' assignment costs. The method still provides good solutions within a short computation time.

4.4. Characteristics of the optimal plan

This section investigates the characteristics of the optimal plan. For this purpose, we consider the optimal plan determined by method M6 for one problem instance with $\tau = 75\%$ for 12 identical ORs and 12 non-identical ORs (with $R = €\,400$). No patient is excluded in this example. Figure 1 and Figure 2 present the following information: (i) the demand D_t which is the total operating time of elective patients with the earliest period equal to t divided by the number S of ORs; (ii) the minimum and the maximum workload of elective patients for different ORs in each period. Recall that the demand for emergency surgery is on average 2 hours for each OR-day and the regular capacity for each OR-day is 8 hours. It can be observed that overtime is scheduled for days 1 and 2 due to a large demand for elective surgery on day 1. Furthermore the workload decreases over time. Another interesting observation is that for the problem with identical ORs the difference between the minimum and maximum OR workloads is small. This implies that the different ORs have a well-balanced workload in each period. However, for the problem with non-identical ORs the workload is not so well balanced in each period. The solution cost for the problem with non-identical ORs is 1.5% larger than the cost to solve the identical ORs problem.

Fig. 1 The workload distribution for a problem with 12 identical ORs.

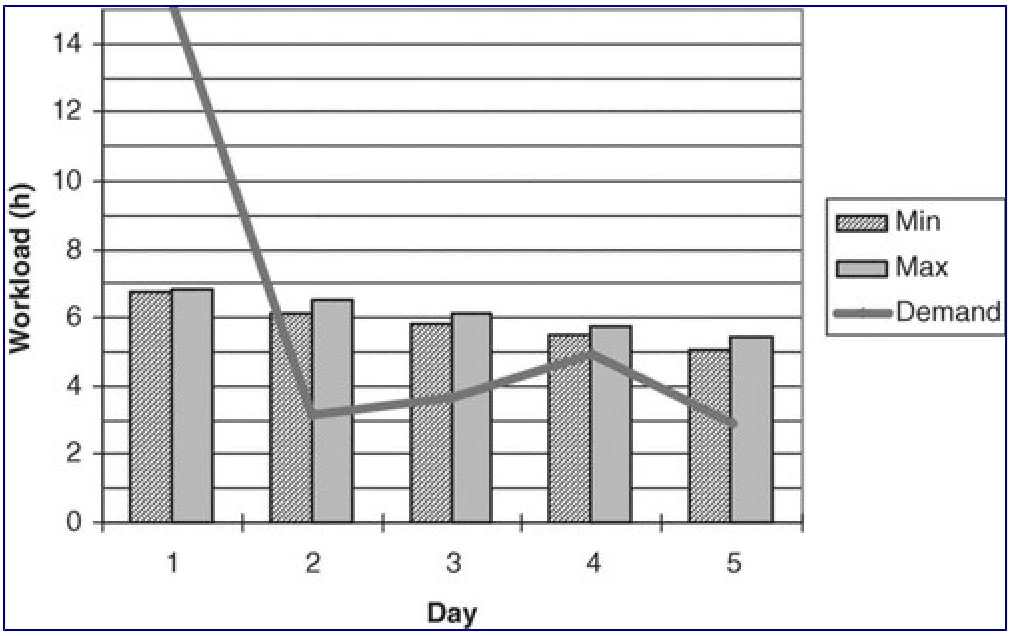
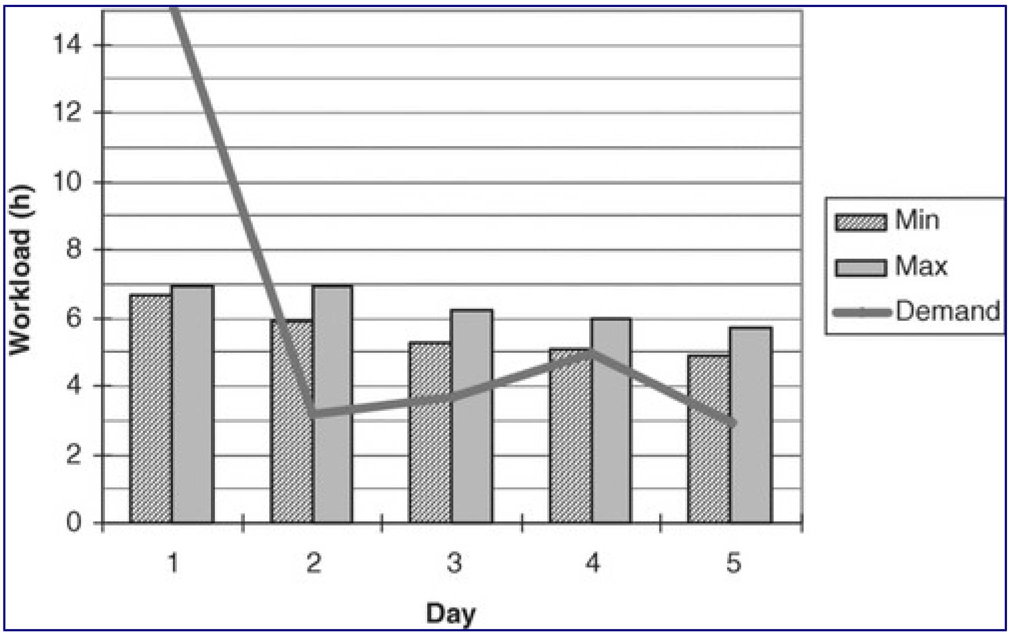


Fig. 2 The workload distribution for a problem with 12 non-identical ORs (R= 400).



5. Conclusions and topics for further research

In this paper we have proposed a stochastic model for planning elective surgeries under uncertain demand for emergency surgery. The planning problem consists in assigning elective patients to different ORs over a planning horizon in order to minimize elective-patient-related costs and the expected utilization costs of ORs related both to underutilization and overutilization.

The problem was formulated as a stochastic integer program. A column-oriented formulation has been proposed in which each column represents a possible assignment of elective patients to a particular OR in a particular period. A column generation approach has been proposed to solve the problem. The pricing problem, a knapsack problem with random sack size, has been solved efficiently using a dynamic programming method.

This approach has the following advantages: it provides a near-optimal solution, and a lower bound that provides a quantitative optimality measure. Numerical results show that it is able to generate near-optimal plans within a short computation time (less than 2 minutes) for problems with practical sizes (12 ORs and about 210 elective patients).

The following observations can be made with regard to the column generation. First, the provided lower bound is very tight; solutions within 2% of the optimum are found. Second, the first-negative strategy is worse than the classical scheme, i.e., the best-negative approach. However, the all-negative strategy is the best one and leads to a significantly smaller computation time. These results place the column generation approach as a very competitive solution method for the elective patients planning problem. However, there are some issues worthy of futher study; such as the use of a two-phase pricing strategy, stabilization techniques and the development of a branch-and-price algorithm.

As previously stated in the paper, there has been little work performed on OR planning methods that explicitly take into account randomness related to emergency surgery. Comparing with a previous work (Lamiri *et al.*, 2008 Lamiri, M., Xie, X., Dolgui, A. and Grimaud, F. 2008. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, 185: 1026–1037. [CrossRef], [Web of Science ®], [Google Scholar]), this paper proposes a more realistic model which considers the assignment of elective patients to ORs, and a more sophisticated cost structure of ORs' capacity utilization. Furthermore, the column generation approach is now able to solve problems of realistic size in a short computation time.

However, several issues need to be further addressed for real-life application of our approach. Elective patient assignment costs were used to model several practical constraints. The model can be easily extended to explicitly take into account some of these constraints, such as surgeon and/or OR availabilities. Such constraints can be respected when generating the columns, i.e., at the level of the column generation subproblem. Deadlines and hospitalization costs can be easily modeled by using patient assignment costs. Future research is needed on the estimation of the patient-related costs. Some automated costing procedures are required to produce these sets of costs. Extending the model to take into account secondary resources (recovery beds, ward beds, intensive care units…) and estimation of the distribution of emergency capacity requirement for each OR-day are two other important avenues of future research.

Deterministic elective surgery durations were assumed in this work. It is known that, in practice, the duration of an elective surgery case depends on the surgeon, the case complexity and many other factors. The proposed model can take into account randomness related to elective surgery, but the column generation approach of this paper does not apply. Indeed, with this extension, pricing subproblems in the column generation approach become knapsack problems with a random item size and random sack size; a problem without known efficient solution methods. Work is in progress to develop a planning method for operating theaters with random surgery times.

Another avenue of future research is the operating theater planning under reliability or robustness constraints such as the probability of exceeding the regular capacity or the probability of the cancellation of the surgery. Such planning models are more difficult to solve but are easier to use by decision makers.

Biographies

Mehdi Lamiri received a B.S. degree in Industrial Engineering from the Ecole Nationale d'Ingénieurs de Tunis, Tunisia, in 2002 and an M.S degree in Industrial and Systems Engineering from the Ecole Centrale Paris, France, in 2004. He is currently working towards a Ph.D. degree in Industrial Engineering at the Ecole Nationale Supérieures des Mines de Saint-Etienne, France, and is expected to graduate in 2007. His current research interests are in the area of stochastic modeling and optimization with applications to healthcare delivery systems.

Xiaolan Xie is a Professor of Industrial Engineering and Head of the Department of Health Care Systems Operation at the Ecole Nationale Supérieure des Mines de Saint Etienne (EMSE), a 190-year old top-10 French graduate engineering school. He has over 20 years of experience in design, planning and scheduling, supply chain optimization, performance evaluation, and maintenance of manufacturing and healthcare systems. He was a research director at INRIA before joining EMSE. He is the author/co-author of over 200 publications including over 70 journal articles in leading international journals and four books on Petri nets. He has served as a Guest Editor of a 2001 special issue of *International Journal of Production Research* on modeling, specification and analysis of manufacturing systems, a 2001 special issue of *IEEE Transactions on Robotics and Automation* on semiconductor manufacturing systems, a 2004 special issue on modelling of reactive systems for *International Journal of Production Research*, and a 2005 special issue on discrete event system techniques for CIM for the *International Journal of Computer Integrated Manufacturing*. He is an Associate Editor of *IEEE Transactions on Automation Science & Engineering* and *IEEE Transactions on Automatic Control* and was an associate Editor of *IEEE Transactions on Robotics and Automation*. He has extensive experience working with European industries. He is the French-leader for the European Project FP6-IST6 IWARD on swarm robots for health services, for FP6-NoE I*PROMS on intelligent machines and production systems, for the GROWTH-ONE project for the strategic design of supply chain networks, and for the GROWTH thematic network TNEE on extended enterprises.

Shuguang Zhang is an engineer at the Industrial Strategy Department of Schneider Electric, Shanghai, China. He received an M.S. degree in Industrial Engineering from the Institut National Polytechnique de Grenoble, France, in 2006, and a B.S. degree from Shanghai Jiao Tong University, China, in 2005.