A Multi-Stage Stochastic Integer Programming Model for Surgery Planning

Serhat Gul
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, serhat.gul@gatech.edu
Brian T. Denton
Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, bdenton@ncsu.edu
John W. Fowler
Department of Supply Chain Management, Arizona State University, john.fowler@asu.edu

We propose a multi-stage stochastic mixed integer programming formulation for the assignment of surgeries to operating rooms (ORs) over a finite planning horizon. We consider the demand for surgery and the duration of surgery to be random variables. The objective is to minimize three competing criteria: expected cost of surgery cancellations, postponements, and OR overtime. We discuss properties of the model and an implementation of the progressive hedging algorithm to find near optimal surgery schedules. We conduct numerical experiments using data from a large hospital to identify managerial insights related to surgery planning and the avoidance of surgery cancellations. Finally, we compare the progressive hedging algorithm to an easy to implement heuristic for practical problem instances to estimate the value of the stochastic solution.

Key words: surgery planning; scheduling; stochastic programming; progressive hedging; heuristics
History: Prepared on June 27, 2012

1. Introduction

Operating rooms (ORs) are responsible for a large portion of total hospital revenues (HFMA 2005) and costs. Therefore, substantial cost reductions might be achieved through better management of ORs. Typically, a two-phase process is followed to plan for a day of surgery. In the first phase, surgeries are assigned to days. This is often done weeks prior to the day of surgery. In the second phase, surgeries are sequenced and scheduled within ORs, often days prior to the day of surgery. Surgery assignment, sequencing, and scheduling decisions have the potential to influence the cost of overtime and cancellations.

Surgery cancellations result in prolonged hospital stays, delayed perioperative treatments, and repeated preoperative tests and treatments. Cancellations have been found to incur a cost of \$1700 - \$2000 per case (Argo et al. 2009). A recent study indicates that as many as 50% of cancellations can be prevented (Gillen et al. 2009). One way to prevent cancellations is to create surgery plans that carefully consider the uncertainty related to the future.

Designing surgery plans is a complicated task due to the uncertainty in demand for surgery and duration of surgery. The occurrence of urgent and emergent cases is one of the reasons that uncertainty in demand is a significant factor (Gerchak et al. 1996, Zonderland et al. 2010). There also exists a considerable amount of uncertainty in demand for elective surgeries. Thus, the mix of surgeries requested varies from day to day. Combining this with uncertainty in the duration of individual surgeries (Gul et al. 2011) makes the task of creating surgery plans challenging.

In this article, we study the problem of assignment of surgeries into future days and ORs over a finite planning horizon. Decisions in our model include scheduling and rescheduling of surgeries where the latter results from cancellations that may occur on the day of surgery. Cancellations are an important consideration, because they are commonly observed and they significantly influence efficiency and quality of patient care. For example, one study found that the percentage of canceled surgeries ranges between 5% - 20% across institutions in the US (Argo et al. 2009).

We formulate a multi-stage stochastic mixed integer program for surgery planning. We consider three competing criteria in the objective function: expected cost of surgery cancellations, postponements (the number of days between when the surgery is requested and the day it is performed), and OR overtime. We implement a customized version of the progressive hedging algorithm (PHA) to find near optimal surgery plans. We also compare the PHA with a deterministic heuristic which is similar to planning rules likely to be used in practice. We use our model to solve practical instances of the problem based on data from a large medical center. Our results provide insight regarding answers to the following three questions:

1. Which factors have significant impact on the increased likelihood of surgery cancellations?

2. What is the value of considering the randomness in demand and total daily surgery durations when planning surgeries?

3. Which PHA parameters have significant impact on the performance and solution quality of the PHA?

The remainder of this article is organized as follows. In the next section, a brief literature review of surgery planning studies is presented. In Section 3, the decision making process is described and a multi-stage stochastic mixed integer programming model is formulated. In Section 4, our implementation of the PHA is discussed. In Section 5, the experimental study is presented. Finally, concluding remarks are given in Section 6.

2. Literature Review

The literature review is divided into three categories of research. The first category is deterministic models for OR planning. The second category includes articles which consider uncertainties related to the surgery durations, but not demand uncertainty for elective surgeries. Since the demand for elective surgeries over the planning period is assumed to be known in these studies, the models are static, i.e., all decisions are given at the beginning of the planning period. The third category of articles considers uncertain elective surgery demands in the context of dynamic planning.

Among articles in the first category of research, Guinet and Chaabane (2003) used a two-phase approach based on weekly OR planning. Their integer programming model assigns surgeries to ORs and particular time blocks of each day over a finite planning horizon. The objective is to minimize the patient's indirect waiting time, i.e., the time between the procedure and hospitalization date, and OR overtime. Their model also considers equipment constraints and availability of surgeons. Fei et al. (2008, 2009, 2010) proposed an integer programming model for optimal assignment of surgeries to ORs and days to minimize OR overtime and maximize OR utilization. They formulated the problem as a set partitioning model and applied a column generation based heuristic to solve the model.

In the second category of articles, Min and Yih (2010) modeled the problem of allocating surgeries to the blocks reserved for different surgery specialties. They formulated the problem as a two-stage stochastic mixed integer program and used a sample average approximation method to solve the problem. Their model also considers the availability of intensive care unit (ICU) beds during the block assignment phase. The length of stay in an ICU bed and surgery durations are the stochastic parameters in their model. The objective function minimizes patient priority based waiting costs and OR overtime costs.

Lamiri et al. (2008a) solved the problem of assigning elective surgeries to periods over a planning horizon while considering the impact of uncertainty related to emergency case arrivals. They first modeled the problem as a stochastic combinatorial optimization problem and then provided a reformulation in the form of a sample average approximation problem. The authors considered expected overtime costs and patient related costs as the performance measures. The surgery durations are assumed to be deterministic in the study. Lamiri et al. (2008b) extended the model in Lamiri et al. (2008a) by considering the allocation of surgeries to ORs. Lamiri et al. (2009) proposed several heuristics to solve the same problem in Lamiri et al. (2008a) and compared the heuristics' performance with the performance of a Monte Carlo optimization method. Hans et al. (2008) also solved a stochastic OR-to-day allocation problem, where the stochasticity exists due to the uncertainty of the surgery durations. Their objective is to minimize the planned slack time reserved in the ORs each day which can be used by surgeries running longer than expected. The authors consider the trade-off between the OR utilization and OR overtime. The authors found that the surgeries having similar duration variability should be clustered together and assigned to the same OR-day.

In the third category of articles, Gerchak et al. (1996) modeled a surgery planning problem as a stochastic dynamic program. The decision process in their study was defined as follows: Each day new requests for elective and emergency surgeries arise. Surgeries are scheduled to the current or future days and previously scheduled surgeries may be canceled. The objectives include maximizing the expected profit gained by scheduling elective cases, and minimizing the expected overtime and surgery cancellation costs. Zonderland et al. (2010) also considered a dynamic decision process where the days are assigned to blocks of surgeries at the beginning of every week for a variety of urgency levels. The different urgency levels include elective surgeries as well as semi-urgent surgeries that must be scheduled within one or two weeks. Based on a Markov decision process model, the authors provided a planning guideline by taking the costs related to the OR idle time, OR overtime, and cancellation of elective surgeries into consideration.

Our work differs from the studies in the first and second category due to the stochastic dynamic setting for planning the surgeries. The articles most similar to ours are those by Gerchak et al. (1996) and Zonderland et al. (2010), who also consider a dynamic decision-making process. This article differs in the following ways. First, Gerchak et al. (1996) allows same-day scheduling after a request arises for a surgery, however this is not a very realistic representation of many surgery practices. Second, the surgery durations generated in their model are independent from each other and identically distributed. Third, they do not consider OR allocation decisions and other scheduling complexities included in our model.

This article also differs from Zonderland et al. (2010) in a number of ways. First, the authors do not consider the assignment of individual surgeries to days, but rather reserve time slots for elective or semi-urgent surgeries each day. Thus, for example, they do not make a distinction between different types of elective surgeries. Furthermore, they make strict assumptions about the nature of uncertainty including that surgery requests arise according to a Poisson process, and surgery durations are assumed to be exponentially distributed. In contrast, our study makes no special assumptions about the random model parameters and our numerical results are based on real data from a large medical center.

3. Problem Description

The model formulated and discussed in the remainder of this article considers the decisions for the dynamic allocation of surgeries to ORs over a finite planning horizon under uncertainty (see Figure 1). The problem is formulated as a multi-stage stochastic mixed integer program (MSSMIP). At each stage, i.e. day, newly requested surgeries are scheduled to future days; furthermore, some previously scheduled surgeries may be cancelled and subsequently rescheduled to a future stage. In addition to assigning each surgery a day, an available OR is also assigned.

At the beginning of each day, it is assumed that random durations for surgeries are observed for the current day. Thus, after the final schedule is determined for each day, the cumulative duration of the surgeries assigned to the ORs, total amount of OR overtime, and cancellations are determined.

Total expected OR overtime, postponement and cancellation costs are the performance measures considered. To reduce overtime, surgeries might be cancelled and rescheduled into future. However, the number of cancellations must be limited, because it results in surgery cancellation and postponement costs. To reflect this, the model includes a per day cancellation and postponement cost associated with each surgery. Furthermore, we assume there exists a time window within which each surgery must be completed. Decisions are made at each day during the planning horizon. Surgeries may also be scheduled to an additional dummy day at the end of the planning horizon. In case there are surgeries already scheduled into the current planning horizon, the capacities of the corresponding ORs are reduced.

The objective of our model is to minimize the daily cost of overtime, postponement and cancellations at a given day, and the expected daily costs of overtime, postponement

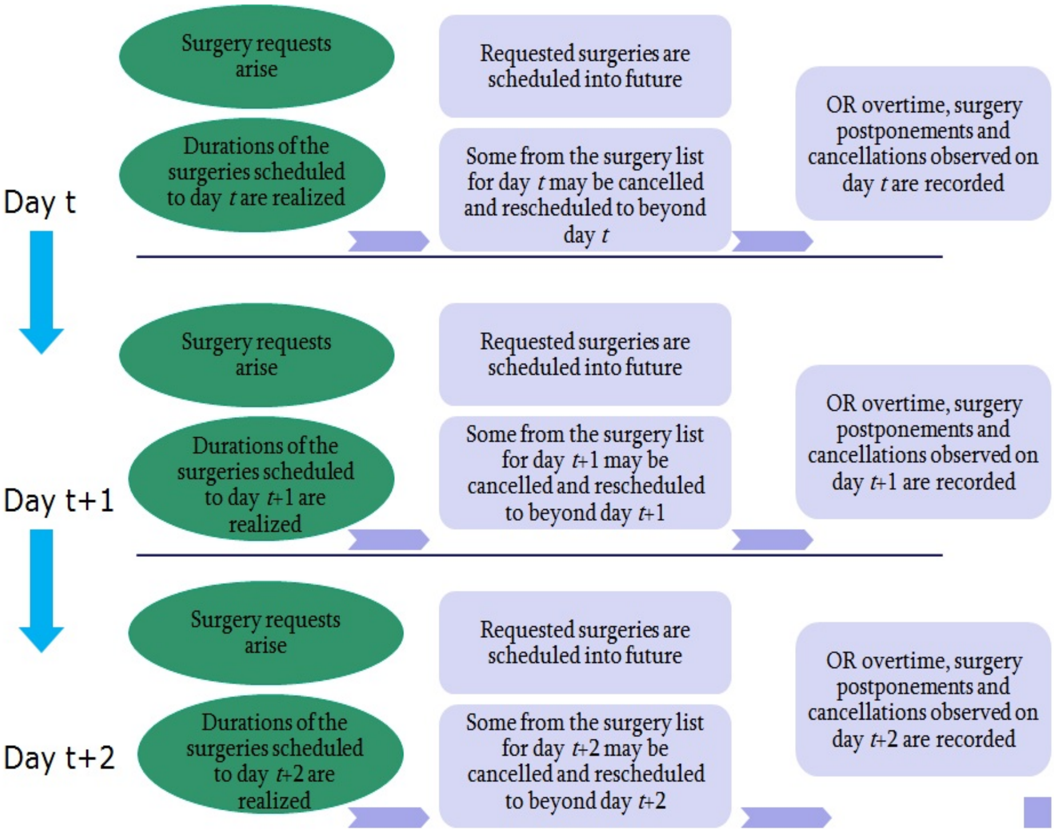

Figure 1 The pattern followed while taking surgery scheduling decisions during a 3-day length of planning period

and cancellations over the future days in the planning horizon. Following is a detailed description of the MSSMIP model.

Indices:

i: surgery index

l,t,u: day index

j: OR index

ω^t: scenario realization at day t

$\omega^{[t]} := (\omega^1, ..., \omega^t)$: history of scenario realization up to day t

Deterministic Parameters:

$$\lambda_{ij} = \begin{cases} 1 & \text{if there is no equipment constraint restricting the assignment} \\ & \text{of surgery } i \text{ to OR } j; \\ 0 & \text{otherwise.} \end{cases}$$

$$p_{iu} = \begin{cases} 1 & \text{if surgery } i \text{ can be assigned to day } u \\ 0 & \text{otherwise.} \end{cases}$$

g_i = lead time (number of days between the earliest day the surgery can be assigned to and the day the request arises) for scheduling surgery i.
h_i = length of time window (number of days between the earliest day and the latest day that the surgery can be assigned to) for scheduling surgery i.

$$a_{ijt} = \begin{cases} 1 & \text{if surgery } i \text{ was already assigned to day } t \text{ and OR } j \text{ before the planning} \\ & \text{horizon starts;} \\ 0 & \text{otherwise.} \end{cases}$$

P_j^t = capacity (in terms of minutes) of OR j at day t
c^i = cost per cancellation of surgery i
l^i = postponement cost per day for surgery i
c^o = OR overtime cost per minute
O = number of ORs
H = length of planning horizon for scheduling surgeries
Random Parameters and Sets:
$d_i(\omega^{[t]})$ = random duration of surgery i under scenario $\omega^{[t]}$
$s(\omega^t)$ = set of surgeries requested at day t according to realization ω^t
$s(\omega^{[t]})$ = set of surgeries requested before and at day t under scenario $\omega^{[t]}$
$\xi^t = (s(\omega^{[t]}), d_i(\omega^{[t]}))$: set of realization history of random parameters at day t

t^{th} *Stage Decision Variables:*

$$x_{iju}^t(\omega^{[t]}) = \begin{cases} 1 & \text{if surgery } i \text{ is assigned to OR } j \text{ and day } u \text{ at day } t \text{ under scenario } \omega^{[t]}; \\ 0 & \text{otherwise,} \end{cases}$$

$$\sigma_{ij}^t(\omega^{[t]}) = \begin{cases} 1 & \text{if surgery } i \text{ from OR } j \text{ is canceled at day } t \text{ under scenario } \omega^{[t]}; \\ 0 & \text{otherwise,} \end{cases}$$

$o_j^t(\omega^{[t]})$ = overtime for OR j observed at day t under scenario $\omega^{[t]}$
$\mathbf{x^t}(\omega^{[t]})$ = vector of values for all decision variables defined for and before day t under scenario $\omega^{[t]}$
The constraint set in the formulation of our problem has a block diagonal structure. There are H blocks of constraints as well as nonnegativity and binary restrictions on the decision variables. Each of the first H-1 blocks contain seven types of constraints, while the last block has only four types. At day H, there does not exist any surgery request since this is the last day of the planning horizon and same-day scheduling decisions are not allowed in the model. Therefore, the model may only give a cancellation decision on this day. Hence, the constraint set is more compact than the previous stages.

For the last day, we have the following formulation for the recourse function, $Q^H(\mathbf{x^{H-1}}(\omega^{[H-1]}), \xi^{H-1}(\omega^{[H-1]}))$.

$$\min \sum_{j=1}^{O}(\sum_{i \in s(\omega^{[H-1]})} c^i \sigma_{ij}^H(\omega^{[H]}) + c^o o_j^H(\omega^{[H]}) + l_i x_{ijH+1}^H(\omega^{[H]})) \tag{3.1}$$

s.t.

$$\sigma_{ij}^H(\omega^{[H]}) - a_{ijH} - \sum_{l=1}^{H-1} x_{ijH}^l(\omega^{[H]}) \leq 0 \qquad \forall\ \omega^{[H]}, i \in s(\omega^{[H-1]}), j \tag{3.2}$$

$$\sum_{j=1}^{O} \sigma_{ij}^H(\omega^{[H]}) - p_{iH+1} \leq 0 \qquad \forall\ \omega^{[H]}, i \in s(\omega^{[H-1]}) \tag{3.3}$$

$$\sum_{i \in s(\omega^{[H-1]})} d_i(\omega^{[H]})(a_{ijH}^0 + \sum_{l=1}^{H-1} x_{ijH}^l(\omega^H) - \sigma_{ij}^H(\omega^{[H]})) - o_j^H(\omega^H) \leq P_j^H \qquad \forall\ j, \omega^{[H]} \tag{3.4}$$

$$x_{ijH+1}^H(\omega^{[H]}), \sigma_{ij}^H(\omega^{[H]}) \in \{0,1\} \quad \forall i; j; \omega^{[H]} \tag{3.5}$$

$$o_j^H(\omega^{[H]}) \geq 0 \quad \forall j; \omega^{[H]} \tag{3.6}$$

The objective function minimizes the postponement, cancellation, and overtime costs at stage *H*. Constraint (3.2) and (3.3) require that a surgery in an OR can only be cancelled on day *H* if it was previously assigned to this day and OR; and if it is possible to reschedule the surgery to another day. Note that the cancellation decision for a surgery can be given more than once over the planning horizon. Constraint (3.4) calculates the overtime for an OR by considering the surgeries scheduled to day *H* but not cancelled. Constraint (3.5) and (3.6) define the nonnegativity and integrality restrictions.

The constraint block for stage t can be regarded as a generic block representing each of the blocks for the previous *H-1* stages. Letting $\mathcal{Q}^{t+1}(\mathbf{x^t}(\omega^{[t]})) = E_{\xi^{t+1}}[Q^{t+1}(x^t(\omega^{[t]}), \xi^{t+1}(\omega^{[t+1]}))]$ for all t, we obtain the following recursion for $Q^t(\mathbf{x^{t-1}}(\omega^{[t-1]}), \xi^t(\omega^{[t]}))$ defined for $t = 2, ..., H-1$.

$$\min \sum_{j=1}^{O}(\sum_{i \in s(\omega^{[t-1]})} c^i \sigma_{ij}^t(\omega^{[t]}) + c^o o_j^t(\omega^{[t]}) + \sum_{i \in s(\omega^{[t]})} \sum_{u=t+1}^{t+g_i+h_i} l_i(u-t) x_{iju}^t(\omega^{[t]})) + \mathcal{Q}^{t+1}(\mathbf{x^t}(\omega^{[t]})) \tag{3.7}$$

s.t.

$$\sum_{u=t+1}^{t+g_i+h_i} \sum_{j=1}^{O} x_{iju}^t(\omega^{[t]}) = 1 \qquad \forall\ \omega^{[t]}, i \in s(\omega^{[t]}) \tag{3.8}$$

$$\sum_{u=t+1}^{t+g_i+h_i} \sum_{j=1}^{O} x_{iju}^t(\omega^{[t]}) = \sum_{j=1}^{O} \sigma_{ij}^t(\omega^{[t]}) \qquad \forall\ \omega^{[t]}, i \in s(\omega^{[t-1]}) \tag{3.9}$$

$$x_{iju}^t(\omega^{[t]}) \leq \lambda_{ij} p_{iu} \qquad \forall\ \omega^{[t]}, i \in s(\omega^{[t]}), j, u = t+1, .., H \tag{3.10}$$

$$\sigma_{ij}^t(\omega^{[t]}) - a_{ijt} - \sum_{l=1}^{t-1} x_{ijt}^l(\omega^{[t]}) \leq 0 \qquad \forall\ \omega^{[t]}, i \in s(\omega^{[t-1]}), j \tag{3.11}$$

$$\sum_{j=1}^{O} \sigma_{ij}^t(\omega^{[t]}) - p_{it+1} \leq 0 \qquad \forall\ \omega^{[t]}, i \in s(\omega^{[t-1]}) \tag{3.12}$$

$$\sum_{i \in s(\omega^{[t-1]})} d_i(\omega^{[t]})(a_{ijt} + \sum_{l=1}^{t-1} x_{ijt}^l(\omega^{[t]}) - \sigma_{ij}^t(\omega^{[t]})) - o_j^t(\omega^{[t]}) \leq P_j^t \qquad \forall\ j, \omega^{[t]} \tag{3.13}$$

$$x^t_{iju}(\omega^{[t]}), \sigma^t_{ij}(\omega^{[t]}) \in \{0,1\} \quad \forall i; j; t = 1, ..., H; u = 2, ..., t + g_i + h_i; \omega^{[t]} \tag{3.14}$$

$$o^t_j(\omega^{[t]}) \geq 0 \quad \forall j; t = 1, ..., H; \omega^{[t]} \tag{3.15}$$

Constraint (3.8) ensures that a surgery must be assigned to an OR in one of the subsequent days after day t if a request arises for this surgery on day t. Constraint (3.9) enforces the assignment of a cancelled surgery to a future day and OR. Constraint (3.10) imposes restrictions on the particular day and OR a surgery may be assigned. When a request arises for a surgery, it must be scheduled within the allowable time window (h_i) for performing the surgery and at least g_i stages into the future. A restriction on the assignment of a surgery to an OR might also exist, defined by constraint (3.10), if the OR does not have all equipment necessary for the surgery. Constraints (3.11) and (3.12) (which are equivalent to (3.2) and (3.3), respectively) define the limits on the decision variables related to the cancellation decisions given on day t. Constraints (3.13) is placed to calculate amount of OR overtime on day t. Constraints (3.14) and (3.15) define the nonnegativity and integrality restrictions on the variables.

The above MSSMIP model is NP-hard. This follows from the fact that an instance of this problem, where the model has only one scenario corresponds to the well known bin packing problem.

4. Solution Methodology

The problem is solved using the progressive hedging algorithm (PHA) proposed by Rockafellar and Wets (1991). The PHA proceeds by applying scenario decomposition to the overall problem iteratively, solving the resulting individual scenario subproblems, and finally aggregating individual scenario solutions. Although the PHA is guaranteed to converge to a global optimal solution asymptotically in the convex case (Rockafellar and Wets 1991), it may converge to only a local optimal solution in this case, because the problem is nonconvex, due to the binary decision variables.

The PHA has been applied in several application areas since the time it was proposed by Rockafellar and Wets (1991) (for example, see Mulvey and Vladimirou (1992) for a financial planning application; Helgason and Wallace (1991) for fisheries management application; Santos et al. (2009) for hydrothermal systems operation planning application). The reader is referred to Wallace and Helgason (1991), Watson and Woodruff (2011) for a detailed discussion about the algorithm implementation.

Many authors of PHA based studies have analyzed the algorithm and proposed ways to improve the overall performance of the PHA based on the special structure of the problem of interest (Mulvey and Vladimirou 1991b,a, Wallace and Helgason 1991, Hvattum and Lokketangen 2009, Watson and Woodruff 2011, Crainic et al. 2011). Background information on our own implementation of the PHA is given in Section 4.2.

4.1. Problem Reformulation

To apply PHA, we begin by reformulating the model to put it in the standard form appropriate for scenario decomposition. Scenario decomposition can be applied when constraints can be separated based on scenarios. In the new formulation, which we refer to as the PHA deterministic equivalent model (PHA-DEM), a new parameter, η, that represents a sequence of consecutive scenarios aggregated over all days (i.e. $\omega^{[1]},\omega^{[2]},...,\omega^{[H]}$) is defined and introduced. Figure 2 illustrates how the reformulation impacts the scenario tree. Figure 2-(a) and Figure 2-(b) compare scenario trees for the MSMIP and PHA-DEM, respectively. Each oval node in the scenario tree represents a particular scenario realization, ω^t, at a particular stage t. The accumulation of oval nodes until stage t (i.e. $\omega^1,\omega^2,...,\omega^t$) defines a particular scenario at day t (i.e. $\omega^{[t]}$). The circle nodes within the oval nodes indicate the surgeries requested at a particular day under the realization that the oval node represents. Note that, for simplicity, the example in Figure 2 assumes that the uncertainty is based only on the surgery requests (i.e. the surgery durations are deterministic).

Figure 2-(a) illustrates that $\omega^{[4]}$ varies based on the scenario represented by $\omega^{[3]}$. The same relation exists also for ($\omega^{[1]}$, $\omega^{[2]}$) and ($\omega^{[2]}$, $\omega^{[3]}$). On the other hand, Figure 2-(b) illustrates an alternative representation of the scenario tree given in Figure 2-(a) where the individual scenarios observed in the particular stages are aggregated over all days to form three scenario sequences, $\eta = 1,2,3$. However, the above redefinition of the scenario tree is not permissible since the solutions found might not be feasible for the overall problem, because they imply decisions that anticipate future uncertain events. Therefore, *nonanticipativity constraints* are required in the PHA-DEM. These constraints enforce the following property: If two scenario sequences, (i.e. $\eta = a,b$), share the same history up to day t, the surgery schedules created progressively over the planning period should always have the same content until day t under the two scenario sequences. In other words, if a decision is given for a surgery at some day l, where $l \leq t$ under scenario sequence a, the

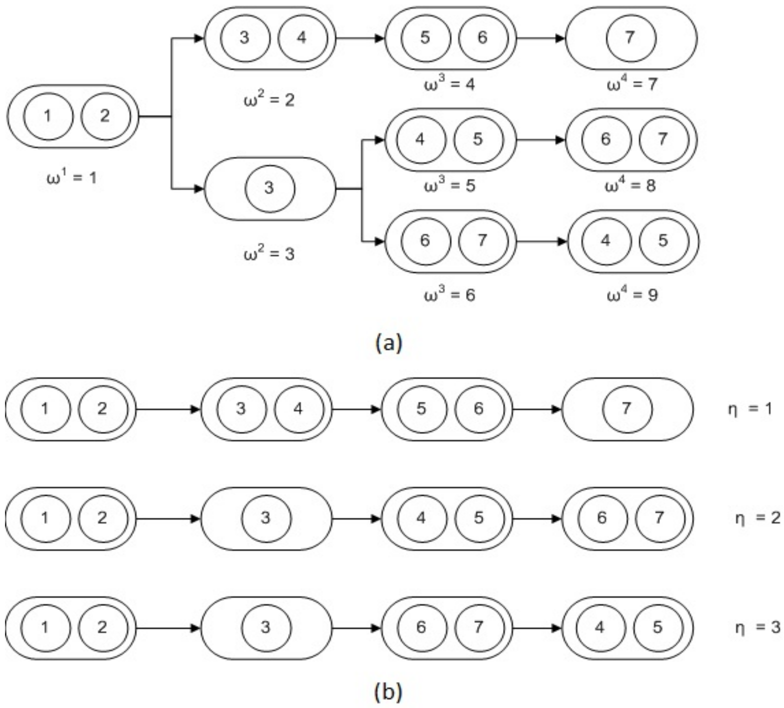

Figure 2 **(a) A scenario tree example illustrating the surgeries that are requested at each day over a four-day planning period (b) The example in (a) is shown in terms of individual scenario sequences**

same decision holds under scenario sequence b.

Following is the additional notation used to formulate the PHA-DEM.

Additional Indices:

Z: number of scenario sequences

N: number of surgeries requested in a sample of scenario tree

η: scenario sequence index

$B(\eta,t)$: scenario bundle index of the surgeries considered for scheduling at stage t under scenario sequence η

Additional Parameters:

$$s^t_{i\eta} = \begin{cases} 1 & \text{if surgery i is requested at day t under scenario sequence } \eta; \\ 0 & \text{otherwise.} \end{cases}$$

$$p_{i\eta u} = \begin{cases} 1 & \text{if surgery i can be assigned to day u under scenario sequence } \eta; \\ 0 & \text{otherwise.} \end{cases}$$

$d_{i\eta}$ = duration of surgery i under scenario sequence η

Pr_η = probability of the occurrence of scenario sequence η

Revised Decision Variables:

$$x^t_{i\eta ju} = \begin{cases} 1 & \text{if surgery i is assigned to OR j and day u at day t under} \\ & \text{scenario sequence } \eta; \\ 0 & \text{otherwise,} \end{cases}$$

$$\sigma^t_{i\eta j} = \begin{cases} 1 & \text{if surgery i from OR j is canceled at day t under scenario sequence } \eta; \\ 0 & \text{otherwise,} \end{cases}$$

$o^t_{\eta j}$ = resulting overtime amount for OR j on day t under scenario sequence η

Additional Decision Variables:

$$x^{B(\eta,t)}_{iju} = \begin{cases} 1 & \text{if surgery i is assigned to day u and OR j at all day-scenario sequence} \\ & \text{combinations in the bundle, B}(\eta\text{,t}), \text{ that day t-scenario } \eta \text{ belongs to;} \\ 0 & \text{otherwise,} \end{cases}$$

The nonanticipativity constraints are also referred to as *bundle constraints*. If the scenario sequences a and b share the same history up to day t, then this indicates they also share the same *scenario bundle* on day t: $B(a,t) = B(b,t)$. Thus, the scheduling decisions given on this day are the same among all scenario sequences placed in the same scenario bundle.

Figure 3 illustrates the scenario bundle concept using the example given in Figure 2. The rectangles covering the oval nodes represent the particular scenario bundles that exist in the example. Since all three scenario sequences have the same realization (e.g. $\omega^1 = 1$) at day 1, $\eta = 1,2,3$ share the same bundle at this day, thus this yields the following equation: $B(1,1) = B(2,1) = B(3,1) = 1$. The second day also contains one scenario bundle, because $\eta = 2$ and $\eta = 3$ share the same history by day 2.

We now show how decisions are synchronized using the bundle constraints. First, recall that the model includes decisions for two different cases: (i) a request arises for a new surgery; (ii) one of the surgeries is canceled. In day 1, for all η, surgeries 1 and 2 are

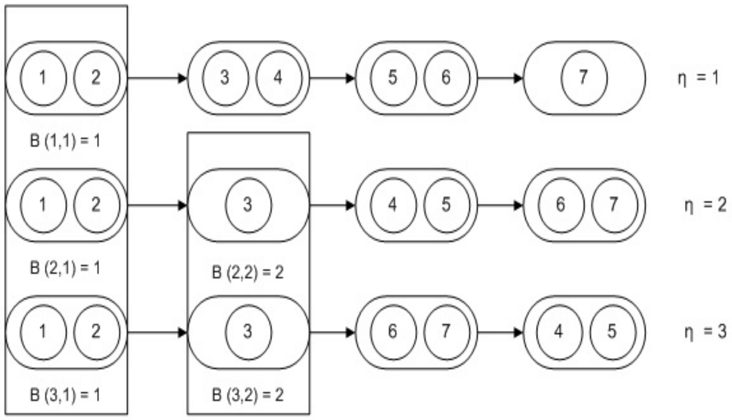

Figure 3 **Representation of scenario bundles by rectangles covering the scenario realizations at a particular day.**

scheduled into the future, corresponding to case (i). One can enforce nonanticipativity using the following constraints:

$$x^1_{i1ju} = x^1_{i2ju} = x^1_{i3ju} \quad \forall j, u = 2,3,4,5 \text{ and } i = 1,2.$$

Similarly, nonanticipativity related to surgery 3 can be generated using:

$$x^2_{32ju} = x^2_{33ju} \quad \forall j, u = 3,4,5.$$

For case (ii), the decisions to reschedule cancelled surgeries under scenario sequences $\eta = 2,3$ are bundled as follows:

$$x^2_{i2ju} = x^2_{i3ju} \quad \forall j, u = 3,4,5 \text{ and } i = 1,2.$$

To facilitate the generation of a separable stochastic program, a new decision variable, called the consensus variable, $x^{B(r,t)}_{iju}$, is defined. The PHA-DEM is formulated as follows:

$$\min \sum_{\eta=1}^{Z} Pr_\eta (\sum_{t=1}^{H} \sum_{j=1}^{O} (c^o o^t_{\eta j} + \sum_{i=1}^{N} c^i \sigma^t_{i\eta j} + \sum_{i=1}^{N} \sum_{u=t+1}^{t+g_i+h_i} l_i(u-t) x^t_{i\eta ju})) \tag{4.1}$$

s.t.

$$x^t_{i\eta ju} = x^{B(\eta,t)}_{iju} \qquad \forall i, \eta, j, t, u > t \tag{4.2}$$

$$\sum_{u=t+1}^{t+g_i+h_i} \sum_{j=1}^{O} x^t_{i\eta ju} = s^t_{i\eta} + \sum_{j=1}^{O} \sigma^t_{i\eta j} \qquad \forall i, \eta, t \tag{4.3}$$

$$x^t_{i\eta ju} \leq \lambda_{ij} p_{i\eta u} \qquad \forall\ i, \eta, j, t, u > t \tag{4.4}$$

$$\sigma^t_{i\eta j} - a_{ijt} - \sum_{l=1}^{t-1} x^l_{i\eta jt} \leq 0 \qquad \forall\ i, r, j, t \tag{4.5}$$

$$\sum_{j=1}^{O} \sigma^t_{irj} - p_{irt+1} \leq 0 \qquad \forall\ i, r, t \tag{4.6}$$

$$\sum_{i=1}^{N} d_{i\eta}(a^0_{ijt} + \sum_{l=1}^{t-1} x^l_{i\eta jt} - \sigma^t_{i\eta j}) - o^t_{\eta j} \leq P^t_j \qquad \forall\ j, \eta \tag{4.7}$$

$$x^t_{i\eta ju}, x^{B(\eta,t)}_{i\eta ju}, \sigma^t_{i\eta j} \in \{0,1\} \quad o^t_{\eta j} \geq 0 \quad \forall i, \eta, j, t, u > t \tag{4.8}$$

The objective function (4.1) is the weighted sum of the total scenario costs over all scenarios. The total scenario cost is weighted by the probability associated with the scenario, Pr_η. The total cost for a scenario includes the total OR overtime cost and surgery cancellation and postponement cost over all days.

Constraint (4.2) is the bundle constraint. Constraints ((4.3)-(4.7)) have the same meaning as ((3.8)-(3.13)) in MSSMIP but using one less constraint. The number of constraints is one less because we are now able to define the parameter, $s^t_{i\eta}$, that indicates whether a surgery is requested or not on a particular day. This parameter helps us formulate both scheduling and rescheduling decisions in one constraint instead of two. Constraint (4.3) sets the conditions to be satisfied to give a scheduling decision at a particular day. Constraint (4.4) defines the allowable days and ORs for the assignment of a particular surgery. Constraints (4.5) and (4.6) together ensure that the cancellation decision for a surgery in an OR can only be made if the surgery was assigned to the OR and day, and if the surgery is allowed to be postponed, respectively. Constraint (4.7) measures overtime values for each OR, each day. Constraint (4.8) defines the nonnegativity and binary restrictions on the decision variables.

Using the PHA-DEM formulation, an augmented Lagrangian relaxation technique is applied by dualizing the bundle constraint. The relaxed formulation still includes the constraints ((4.3)- (4.8)) in the constraint set. However, the objective function (4.1) is now:

$$\min \sum_{\eta=1}^{Z} Pr_\eta (\sum_{t=1}^{H} \sum_{j=1}^{O} (c^o o^t_{\eta j} + \sum_{i=1}^{N} c^i \sigma^t_{i\eta j} + \sum_{i=1}^{N} \sum_{u=t+1}^{t+g_i+h_i} l_i(u-t) x^t_{i\eta ju})$$

$$+\sum_{i=1}^{N}\sum_{t=1}^{H}\sum_{j=1}^{O}\sum_{u=t+1}^{t+g_i+h_i}\mu_{i\eta ju}^{t}(x_{i\eta ju}^{t}-x_{iju}^{B(\eta,t)})+\frac{\rho}{2}\sum_{i=1}^{N}\sum_{t=1}^{H}\sum_{j=1}^{O}\sum_{u=t+1}^{t+g_i+h_i}\|x_{i\eta ju}^{t}-x_{iju}^{B(\eta,t)}\|^2) \quad (4.9)$$

where $\mu_{i\eta ju}^{t}, \forall i,\eta,t,j,u$ denote the Lagrangian multipliers; ρ is the penalty parameter; and $\|.\|$ is the ordinary Euclidean norm. The additional components in the function (4.9) penalizes the violation of the bundle constraint. Since $x_{i\eta ju}^{t}, x_{iju}^{B(\eta,t)} \in \{0,1\}$, the penalty component in (4.9) is rewritten as follows:

$$\|x_{i\eta ju}^{t}-x_{iju}^{B(\eta,t)}\|^2 = x_{i\eta ju}^{t}-2x_{i\eta ju}^{t}x_{iju}^{B(\eta,t)}+x_{iju}^{B(\eta,t)} \quad (4.10)$$

To make the deterministic equivalent formulation scenario separable requires fixing the consensus variable, $x_{iju}^{B(\eta,t)}$. Using a *proximal point method* (Rockafellar 1976), this value can be estimated using the weighted sum calculation:

$$\hat{x}_{iju}^{B(\eta,t)} = \sum_{\eta\in B(\eta,t)}^{Z}\frac{Pr_\eta}{\sum_{\eta\in B(\eta,t)}Pr_\eta}x_{i\eta ju}^{t} \quad \forall i,\eta,t,j,u. \quad (4.11)$$

Note that (4.10) does not contain a quadratic term anymore after replacing $x_{iju}^{B(\eta,t)}$ with its estimation, $\hat{x}_{iju}^{B(\eta,t)}$, which facilitates the solution of the subproblems following the scenario decomposition.

Equation (4.11) calculates the weighted sum of the individual scheduling decision variables within a decision bundle. The weights are set by normalizing the probability of the scenario associated with a decision variable. The complete formulation, which we refer to as the separable deterministiu equivalent model (PHA-SDEM) is:

$$\min\sum_{\eta=1}^{Z}Pr_\eta(\sum_{t=1}^{H}\sum_{j=1}^{O}(c^o o_{\eta j}^{t}+\sum_{i=1}^{N}c^i\sigma_{i\eta j}^{t}+\sum_{i=1}^{N}\sum_{u=t+1}^{t+g_i+h_i}l_i(u-t)x_{i\eta ju}^{t})$$
$$+\sum_{i=1}^{N}\sum_{t=1}^{H}\sum_{j=1}^{O}\sum_{u=t+1}^{t+g_i+h_i}\mu_{i\eta ju}^{t}(x_{i\eta ju}^{t}-\hat{x}_{iju}^{B(\eta,t)})+\frac{\rho}{2}\sum_{i=1}^{N}\sum_{t=1}^{H}\sum_{j=1}^{O}\sum_{u=t+1}^{t+g_i+h_i}(x_{i\eta ju}^{t}-2x_{i\eta ju}^{t}\hat{x}_{iju}^{B(\eta,t)}))$$
$$(4.12)$$

s.t.

$$\sum_{u=t+1}^{t+g_i+h_i}\sum_{j=1}^{O}x_{i\eta ju}^{t} = s_{i\eta}^{t}+\sum_{j=1}^{O}\sigma_{i\eta j}^{t} \qquad \forall i,\eta,t \quad (4.13)$$

$$x^t_{i\eta ju} \le \lambda_{ij} p_{i\eta u} \qquad \forall\ i,\eta,j,t,u>t \tag{4.14}$$

$$\sigma^t_{i\eta j} - a_{ijt} - \sum_{l=1}^{t-1} x^l_{i\eta jt} \le 0 \qquad \forall\ i,r,j,t \tag{4.15}$$

$$\sum_{j=1}^{O} \sigma^t_{irj} - p_{irt+1} \le 0 \qquad \forall\ i,r,t \tag{4.16}$$

$$\sum_{i=1}^{N} d_{i\eta}(a^0_{ijt} + \sum_{l=1}^{t-1} x^l_{i\eta jt} - \sigma^t_{i\eta j}) - o^t_{\eta j} \le P^t_j \qquad \forall\ j,\eta \tag{4.17}$$

$$x^t_{i\eta ju}, \sigma^t_{i\eta j} \in \{0,1\} \quad o^t_{\eta j} \ge 0 \quad \forall i,\eta,j,t,u>t. \tag{4.18}$$

Note that the last term of (4.10) is ignored in the PHA-SDEM objective function, (4.12), because it is fixed. Constraints ((4.13)-(4.17)) define the same feasible space as the constraints ((4.3)-(4.7)).

Note that the consensus variable in PHA-DEM is represented by its estimation in PHA-SDEM, $\hat{x}^{B(\eta,t)}_{iju}$, which is called a *consensus parameter*. The consensus parameter is also an estimation of the implementable solution at a given iteration of the PHA. However, there is no guarantee that the estimated implementable solution would be a feasible solution for PHA-DEM. If this solution is also feasible in PHA-SDEM, then it is called an *admissible solution*. The goal of the PHA is to identify a good solution (ideally the optimal solution) among all admissible and implementable solutions.

The mixed integer programming formulation for a particular scenario subproblem model (PHA-SSM) is given as:

$$\min \sum_{t=1}^{H}\sum_{j=1}^{O}(c^o o^t_{\eta j} + \sum_{i=1}^{N} c^i \sigma^t_{i\eta j} + \sum_{i=1}^{N}\sum_{u=t+1}^{t+g_i+h_i} l_i(u-t)x^t_{i\eta ju}) + \sum_{i=1}^{N}\sum_{t=1}^{H}\sum_{j=1}^{O}\sum_{u=t+1}^{t+g_i+h_i} \mu^t_{i\eta ju}(x^t_{i\eta ju} - \hat{x}^{B(\eta,t)}_{iju})$$
$$+ \frac{\rho}{2}\sum_{i=1}^{N}\sum_{t=1}^{H}\sum_{j=1}^{O}\sum_{u=t+1}^{t+g_i+h_i}(x^t_{i\eta ju} - 2x^t_{i\eta ju}\hat{x}^{B(\eta,t)}_{iju}) \tag{4.19}$$

s.t.

$$\sum_{u=t+1}^{t+g_i+h_i}\sum_{j=1}^{O} x^t_{i\eta ju} = s^t_{i\eta} + \sum_{j=1}^{O}\sigma^t_{i\eta j} \qquad \forall i,t \tag{4.20}$$

$$x^t_{i\eta ju} \le \lambda_{ij} p_{i\eta u} \qquad \forall\ i,j,t,u>t \tag{4.21}$$

$$\sigma^t_{i\eta j} - a_{ijt} - \sum_{l=1}^{t-1} x^l_{i\eta jt} \le 0 \qquad \forall\ i,\eta,j,t \tag{4.22}$$

$$\sum_{j=1}^{O} \sigma_{i\eta j}^{t} - p_{i\eta t+1} \leq 0 \qquad \forall\ i,\eta,t \tag{4.23}$$

$$\sum_{i=1}^{N} d_{i\eta}(a_{ijt} + \sum_{l=1}^{t-1} x_{i\eta jt}^{l} - \sigma_{i\eta j}^{t}) - o_{\eta j}^{t} \leq P_{j}^{t} \qquad \forall\ j \tag{4.24}$$

$$x_{i\eta ju}^{t}, \sigma_{i\eta j}^{t} \in \{0,1\} \quad o_{\eta j}^{t} \geq 0 \quad \forall i,j,t,u > t \tag{4.25}$$

The objective function (4.19) corresponds to one of the scenario costs which are aggregated in the objective function (4.12) of the PHA-SDEM. Constraint set ((4.20)-(4.25)) is also a subset of the constraint set ((4.13)-(4.18)) which should be satisfied for all scenarios rather than only for one scenario.

4.2. Progressive Hedging Algorithm

In this section, we describe our implementation of PHA. Let k denote the index for the iteration number of the PHA, $\mu_{i\eta ju}^{t(k)}\ \forall i,\eta,t,j,u$ the Lagrangian multipliers and ρ^{k} the penalty parameters at iteration k. Then, the general steps of the PHA are stated as follows:

PHA

Set algorithm terminates = false, $k=1, \rho^{(k)}=0, \mu_{i\eta ju}^{t(k)}=0\ \forall i,\eta,t,j,u$
while algorithm terminates = false
  for $\eta = 1$ to Z
    Solve the PHA-SSM to obtain $x_{i\eta ju}^{t(k)}\ \forall i,\eta,t,j,u$
  end for
  Calculate the consensus parameter: $\hat{x}_{iju}^{B(\eta,t)}\ \forall i,\eta,t,j,u$
  if $k > 1$
    Update the penalty parameter according to the following scheme:
    $\rho^{(k+1)} = \alpha\rho^{(k)}$, where $\alpha > 0$
  end if
  Update the Lagrangian multipliers according to the following scheme:
  $\mu_{i\eta ju}^{t(k+1)} = \mu_{i\eta ju}^{t(k)} + \rho^{(k)}(x_{i\eta ju}^{t(k)} - \hat{x}_{iju}^{B(\eta,t)})$
  if $x_{i\eta ju}^{t(k)} = \hat{x}_{iju}^{B(\eta,t)}\ \forall i,\eta,t,j,u$
    Set algorithm terminates = true

```
    end if
    else
        Set k = k + 1
    end else
end while
```

4.3. Enhanced Progressive Hedging Algorithm

In our implementation, we took advantage of the special structure of the model formulation to accelerate the computational performance of the PHA and improve the quality of the PHA solutions. We refer to this algorithm as the *enhanced progressive hedging algorithm* (EPHA). The degree of violation of the bundle constraints and decisions taken by the majority of the variables in the decision bundles motivate the Lagrangian multiplier update method. We also analyze if a penalty update method utilizing the information about the convergence pattern of the primal and dual variables may enhance the solutions.

In the PHA literature, there are many other studies that propose enhancements on the PHA based on the special structures of the models. We first present a brief review of enhancements proposed for various problems. Then, we discuss the methods of our EPHA.

4.3.1. PHA enhancements. Mulvey and Vladimirou (1991b,a) discussed the trade-off between the selection of high and low values for the penalty parameters and the impact of the problem structure into this selection. They also discussed the benefits of the dynamic penalty adjustment methods. Helgason and Wallace (1991), Listes and Dekker (2005) discussed the sensitivity of the convergence of the PHA to the choice of penalty parameter. Hvattum and Lokketangen (2009) proposed a method to set a direction of improvement while updating the penalty parameters. They tested the case where there exists parameters for individual nonanticipativity constraints in the model. Watson and Woodruff (2011) also proposed methods to set the penalty parameters for individual nonanticipativity constraints for a class of resource allocation problems.

In our experiments, we observe that updating penalty parameters based on the information on the convergence pattern in the primal and dual space did not achieve significantly better results than keeping the penalty parameter constant. Our experiments suggest that the initial values of the Lagrangian multipliers have significant impact on the quality of the final solution. We selected the initial values after assessing the trade-off between the marginal improvement in solution quality and additional computational time needed as a result of a variation in the values. Mulvey and Vladimirou (1991a), Santos et al. (2009) also discussed the importance of the initial estimates for the Lagrangian multipliers and tested warm start methods including simple heuristics to find reasonable initial values.

It is well known that in the non-convex case, the PHA is not guaranteed to converge (Takriti and Birge 2000). Watson and Woodruff (2011) defined some techniques to detect non-convergence in the form of cyclic behaviors. Whenever they detect a cycle for a variable, they fix the variable value using a simple rule (the largest value of the variable across scenarios is selected). In our case, the Lagrangian multiplier update method prevents cycling. Since the Lagrangian multipliers are defined only for binary variables, the method aims to favor one feasible value over the other primarily based on the selection in the majority of the subproblems. When the majority is not achieved, the postponement and cancellation costs are also considered.

Due to the typically large number of subproblems to be solved following the scenario decomposition at each PHA iteration, computational efficiency in subproblems is important. Furthermore, it has been shown that the PHA is often a reasonable heuristic to use if there exists an efficient algorithm to solve the subproblems of a very large scale stochastic mixed integer problem (Watson and Woodruff 2011). Takriti et al. (1996) developed methods to solve the subproblems of their multi-stage stochastic production planning problem. Barro and Canestrelli (2005) further decomposed the subproblems of a dynamic portfolio management problem into stages to solve those efficiently. An important reason which necessitates the implementation of an efficient solution method on the subproblems is that each subproblem has a quadratic objective function due to the penalty component. Haugen et al. (2001) relaxed the quadratic term in the subproblem objective function and applied a dynamic programming approach to find an optimal solution for the relaxed subproblems. Listes and Dekker (2005) solved the linear relaxation of the subproblems of a robust airline fleet composition problem, which contained integer variables, and used a simple rounding procedure to find a feasible solution for the overall problem. In our problem formulation, the quadratic component is linearized since the corresponding component includes binary variables.

4.3.2. EPHA Implementation. In this subsection, we present our penalty update and Lagrangian multiplier update methods, and EPHA termination criterion.

4.3.2.1. *Penalty parameter setting and update.* We set a constant value for the penalty parameter after conducting some experimental analysis. The experimental analysis was based on the observation of the trade-off between fast convergence to a suboptimal solution (when ρ is too large) and slow convergence to a near optimal solution in the primal feasible space (when ρ is too low).

Next, the method proposed in Hvattum and Lokketangen (2009) was used to compare the convergence rate at iterations k and $k-1$, increasing ρ if it appears that the convergence rate is decreasing. ρ is decreased if the current status is closer to consensus among variables at iteration $k-1$ than at iteration k. Let $\Delta_D^{(k)}$ and $\Delta_P^{(k)}$ be indicators of the convergence rates in the dual space and in the primal space, respectively. Let b index a unique bundle among the ones represented by all $B(\eta,t)$'s, and B represent the total number of unique bundles. Then, equations ((4.26) – (4.27)) define the penalty update method as follows:

$$\Delta_P^{(k)} = \sum_{i=1}^{N}\sum_{b=1}^{B}\sum_{j=1}^{O}\sum_{u=t+1}^{H+1}(\hat{x}_{iju}^{b(k)} - \hat{x}_{iju}^{b(k-1)})^2 \tag{4.26}$$

$$\Delta_D^{(k)} = \sum_{i=1}^{N}\sum_{\eta=1}^{Z}\sum_{t=1}^{H}\sum_{j=1}^{O}\sum_{u=t+1}^{H+1}(x_{i\eta ju}^{t(k)} - \hat{x}_{iju}^{B(\eta,t)(k)})^2 \tag{4.27}$$

$$\rho^{(k+1)} = \begin{cases} \delta_D \rho^{(k)} & \text{if} \Delta_D^{(k)} - \Delta_D^{(k-1)} > 0 \\ \frac{1}{\delta_P}\rho^{(k)} & \text{else if} \Delta_P^{(k)} - \Delta_P^{(k-1)} > 0 \end{cases} \tag{4.28}$$

where $\delta_P > 1$ and $\delta_D > 1$ in (4.28) are fixed multipliers.

4.3.2.2. *Lagrangian multiplier update.* We use a Lagrangian multiplier update method that ensures convergence of the algorithm. The method aims to achieve convergence of the consensus parameter value to one of the two feasible values: 0 or 1 (see Crainic et al. (2011) for a similar approach). The selection among these two values as the convergence point is made according to the majority of the variable values in a bundle. We define a threshold parameter called θ to help define the majority condition. Once the value of the consensus parameter is greater than θ, the majority is assumed to be achieved. In the case that no value is favored by the majority (i.e consensus parameter value is equal to θ), the value to which the consensus parameter should converge is determined according to the cancellation and postponement costs.

The approach is based on the following observation. If $\hat{x}_{iju}^{B(\eta,t)(k)}$ is greater than θ, then this indicates that the majority of the scenario subproblem solutions within the associated bundle dictate the assignment of surgery i to OR j on day u. Our method decreases the values of the Lagrangian multipliers in the subproblems where the relevant variable value is 0. The expectation here is that the surgery is assigned to the same day and OR in the following iterations. On the other hand, if $\hat{x}_{iju}^{B(\eta,t)(k)}$ is less than θ, this shows that surgery i is not assigned to OR j on day u in the majority of the subproblems. Then, the Lagrangian multiplier values are increased in the subproblems where the relevant variable is equal to 1. In case $\hat{x}_{iju}^{B(\eta,t)(k)}$ is equal to θ, this means no particular day-OR couple is favored for the assignment of surgery i among subproblems. If the same day is selected in all subproblems, then any OR is favored for the assignment. Otherwise, the day to be favored is selected according to the cancellation and postponement costs. If the cancellation cost is greater than the postponement cost, then the latest feasible day is preferred. Therefore, the values of the Lagrangian multipliers are increased for the relevant variables that assign surgery i to the earlier days. If the postponement cost is greater than the cancellation cost, then the earliest feasible day is favored to reduce postponements. This requires an increase in the values of the Lagrangian multipliers for the relevant variables that assign surgery i to the later days. Updates are computed as follows:

$$\mu_{i\eta ju}^{t(k+1)} = \begin{cases} \mu_{i\eta ju}^{t(k)} + \rho^{(k)}(x_{i\eta ju}^{t(k)} - \hat{x}_{iju}^{B(\eta,t)(k)}) & \text{if } \hat{x}_{iju}^{B(\eta,t)(k)} < \theta; x_{i\eta ju}^{t(k)} = 1 \\ \mu_{i\eta ju}^{t(k)} - \rho^{(k)}|(x_{i\eta ju}^{t(k)} - \hat{x}_{iju}^{B(\eta,t)(k)})| & \text{if } \hat{x}_{iju}^{B(\eta,t)(k)} > \theta; x_{i\eta ju}^{t(k)} = 0 \\ \mu_{i\eta ju}^{t(k)} + \rho^{(k)}(x_{i\eta ju}^{t(k)} - \hat{x}_{iju}^{B(\eta,t)(k)}) & \text{if } \hat{x}_{iju}^{B(\eta,t)(k)} = \theta; x_{i\eta ju}^{t(k)} = 1 \text{ and} \\ & c^i \geq l_i; u \neq \max\{u : x_{i\eta ju}^{t(k)} = 1, (\eta,t) \in B(\eta,t)\} \\ \mu_{i\eta ju}^{t(k)} + \rho^{(k)}(x_{i\eta ju}^{t(k)} - \hat{x}_{iju}^{B(\eta,t)(k)}) & \text{if } \hat{x}_{iju}^{B(\eta,t)(k)} = \theta; x_{i\eta ju}^{t(k)} = 1 \text{ and} \\ & c^i < l_i; u \neq \min\{u : x_{i\eta ju}^{t(k)} = 1, (\eta,t) \in B(\eta,t)\} \\ \mu_{i\eta ju}^{t(k)} & \text{otherwise,} \end{cases} \quad (4.29)$$

4.3.2.3. *Termination criteria.* The EPHA terminates when the following condition is satisfied:

$$\sum_{\eta=1}^{Z} Pr_\eta \sum_{i=1}^{N} \sum_{t=1}^{H} \sum_{j=1}^{O} \sum_{u=t+1}^{H+1} |x_{i\eta ju}^{t(k)} - \hat{x}_{iju}^{B(\eta,t)(k)}| \leq \epsilon \quad (4.30)$$

This can be interpreted as a measure of the bundle constraint violation being sufficiently small.

4.3.2.4. *Other considerations.* It is possible that the objective function coefficients of a subproblem may not change from one iteration to another due to the method we propose for updating Lagrangian multipliers. The Lagrangian multiplier for a decision variable is not updated at an iteration in case the value of the variable is equal to the value taken by the sufficient majority of the variables in the bundle. Therefore, the objective function coefficients of a subproblem may remain the same if none of its variables require an update for the Lagrangian multipliers. We detect those subproblems at each iteration to minimize the number of times that the subproblem solution routine is called.

5. Experimental Study

In our experiments, we address the three research questions discussed in Section 1. We address the questions, separately, in the following subsections. We tested our methods using the data from an Outpatient Procedure Center at Mayo Clinic from the year 2006 for 4034 patients (Gul et al. 2011). We generated scenario trees representing arrivals of surgery requests over a planning period. Each problem instance is based on a particular service including: Urology, Ophthalmology, Pain Medicine, and Oral Maxillofacial. Surgeries of a service are grouped into acuity levels. Urology and Pain Medicine have 5 acuity levels, while Ophthalmology and Oral Maxillofacial have 2. The probability distribution for the duration of a surgery is based on its acuity level.

We conducted experiments with moderate-size test cases that can be solved to optimality. We compared generated solutions with the optimal solutions to evaluate the EPHA. We also conducted experiments to estimate the value of the stochastic solution (VSS) by comparing the EPHA with a deterministic heuristic for a set of large test cases. The moderate-size case included, on average, 27 surgeries to be scheduled during a 6-day planning period. These cases represent the surgery scheduling process during a typical week (i.e. 5 days). The surgeries that can not be scheduled for the current week are assigned to a dummy day at the end of the planning period. The large test cases consider 50 surgeries, on average, to be scheduled during an 11-day planning period. These cases represent a bi-weekly surgery scheduling process (i.e. 10 days). A dummy day is also included for the surgeries that can not be performed within the two weeks considered.

In our test cases, two ORs are open every day for each surgical service. Each OR operates for 8 hours, daily. Note that the planned overtime amount is determined to provide realistic performance measure values for the scenario trees tested.

Each experiment is performed on a scenario set that consists of 20 different scenario trees of the same size. We report the average and worst-case performance measure values. The EPHA algorithm was implemented in Microsoft Visual C++ 2008 using CPLEX 12 Concert Technology. The experiments were conducted on an Intel Core i5 PC with processors running at 2.27 GHz and 4 GB memory under Windows XP.

5.1. Generating Problem Instances

We generated a scenario tree for each problem instance. There are two parameters that determine the size of the tree: (1) n_s, *number of stages* and (2) n_o, *number of different outcomes observed at each stage except the first stage.* The realization at stage 1 is assumed to be known, thus there is only one outcome at this stage. There exists $n_o^{n_s-1}$ scenario sequences in total in a scenario tree. The generated scenario outcome at stage t does not depend on the past outcomes. The use of independently generated scenario outcomes across stages allows us to use the *common samples* approach (Chiralaksanakul and Morton 2003). The common samples approach indicates that the generated outcomes may exist more than once and the same number of times at a particular stage of a scenario tree. In particular, among the n_o^{t-1} outcomes at stage t, where $t > 1$, in our scenario tree, only n_o of them are unique.

A scenario outcome is characterized by the number of surgeries requested from each acuity level. For example, assume that we generate a scenario tree for the case where $n_s = 4$ and $n_o = 2$ for the surgical service of Pain Medicine. Let the number of surgery requests for acuity level i, where $i = 1, 2, 3, 4, 5$, be represented by $acu(i)$. Then, an outcome on a particular day can be represented by an $acu(i)$ set. The value of each $acu(i)$ is independently sampled according to the probability distribution fit for the surgery request frequency in the data. A sample tree with its $acu(i)$'s set is illustrated in Figure 4. Each rectangle in the tree denotes a scenario outcome at the corresponding stage. The numbers in the rectangles represent a vector of $acu(i)$ values for $i = 1, 2, 3, 4, 5$.

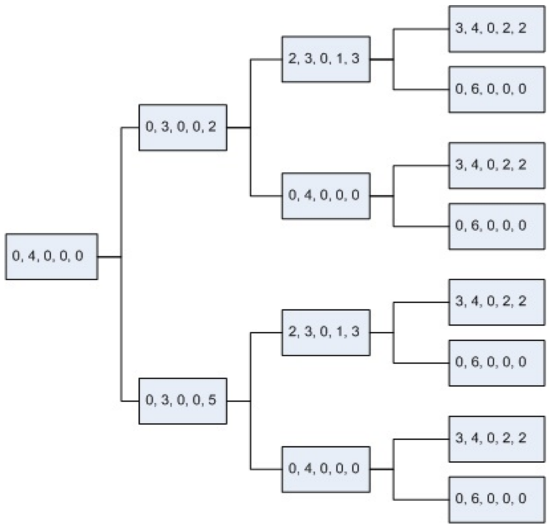

Figure 4 **A sample scenario tree with $n_s = 4$ and $n_o = 2$ represents the requests for urology surgeries over a 4-day planning period. Each index, separated by comma, denotes the number of surgeries requested from a particular acuity level**

5.2. Analysis of the EPHA solutions

In this section, we summarize some of the analysis that was performed for the EPHA.

5.2.1. Sensitivity to EPHA parameters. We first tested the EPHA to identify the significant factors or methods affecting the solution quality and algorithm performance. We investigated whether having the penalty update method achieves improvements over the solutions found by keeping ρ constant (i.e. $\delta_P = 1$, $\delta_D = 1$) through the iterations. We varied the values of δ_P and δ_D keeping everything else constant, and then observed the changes in the average objective function value and computational time. Table 1 shows that the penalty update method improves the computational time (improvement of 8%), while deteriorating the objective value only slightly by 0.5 %.

The Lagrangian multiplier update method presented in equation (4.29) also provides

Table 1 Sensitivity of the objective function value and CPU time to the variations in the penalty update multipliers

(δ_P, δ_D)	Expected cost	CPU time
(1,1)	116708	443.73
(2,2)	117173	443.51
(2,1.5)	117334	467.25
(2,1)	117339	450.95
(4,1)	117185	408.81
(16,1)	117268	422.30

significant improvements to computation time as it prevents cycling and ensures that the algorithm converges.

5.2.2. Comparison of the PHA with the optimal solution. In this section, we compare the PHA solutions with the solutions found after directly solving the DEM using CPLEX 12 for the moderate-size test cases. We report the solution times and objective function values. We report the best solution found within a time limit of 3 hours for cases in which the optimal solution is not achieved.

We compared the solutions on 8 different instance sets (see Table 2). Each set is characterized by the values of three different attributes: Ratio of cancellation to postponement cost, overtime cost, and expected standard deviation of the cumulative duration of the surgeries to be performed each day. The ratio of cancellation to postponement cost can be either greater than or less than 1. The cancellation costs were based on the average cost ($1700-$2000) reported in practice (Argo et al. 2009). The cancellation cost in our experiments was set to either low ($1200/surgery) or high ($2400/surgery). The postponement cost was fixed at $1800. The overtime cost was set to either low ($30/minute) or high

($90/minute). These overtime values are selected to have only one surgery performed during overtime, on average. The mean cumulative duration was set to 540 minutes (9 hours) for the surgeries to be scheduled to an OR that is open for 8 hours during regular time. To consider cases having low and high variability in the surgery requests, the standard deviation of cumulative surgery durations was set to either a low of 78 or a high of 234.

Table 3 compares the EPHA solutions with the CPLEX solutions. Note that the gap

Table 2 Characterization of the instance sets

Instance Set #	Cancellation to postponement cost ratio	Overtime cost	Standard deviation of cumulative durations
1	>1	30	78
2	>1	30	234
3	>1	90	78
4	>1	90	234
5	<1	30	78
6	<1	30	234
7	<1	90	78
8	<1	90	234

between the EPHA and CPLEX solutions is 1% on average. CPLEX found the optimal solution immediately for some instances, while it required a significant amount of time to solve more difficult instances (failing to find the optimal solution in 17% of the cases). The EPHA significantly outperformed CPLEX on the difficult instances.

5.3. Model Sensitivity Analysis

5.3.1. Sensitivity to Cost Parameters. In this section, we analyze the sensitivity of optimal solutions to the changes in the cost coefficients: cancellation, postponement, and OR overtime costs. We emphasize the impact of the changes on the number of surgery cancellations over a planning period. The average number of cancellations and expected total costs are compared in Tables 4 and 5. In the instances presented, the cancellation and postponement costs were varied while keeping the overtime cost constant.

Table 3 Comparison of the PHA solutions with the CPLEX solutions of the deterministic equivalent model (DEM)

	Expected Cost (in dollars)		Worst-Case Cost (in dollars)		CPU time (in seconds)	
Instance Set #	CPLEX	PHA	CPLEX	PHA	CPLEX	PHA
1	83412	83475	93675	93930	0.54	71.59
2	88266	88599	96165	96465	0.50	296.47
3	109295	110529	126765	127605	6513.62	408.38
4	127366	128529	151164	151697	574.04	497.49
5	83427	83563	93765	93765	0.60	62.00
6	88462	89587	96765	97005	0.49	182.93
7	107856	111592	119337	122380	5903.05	574.61
8	122233	124754	145935	146531	1353.59	493.82

Table 4 shows that the expected total cost increases significantly when the postponement cost is increased while the cancellation cost is constant. However, there was only a slight increase in total cost when the cancellation cost was increased while the postponement cost was fixed.

Table 5 reveals that there are two main factors affecting the number of cancellations: (1) the ratio of postponement cost to cancellation cost, (2) the summation of cancellation and postponement costs. Note that when the first factor, the ratio of postponement to cancellation cost is less than 1, then the chance of observing a cancellation is low. If the ratio is greater than or equal to 1 then we observe a significant number of cancellations.

The first factor is important, because if the postponement cost is low, then the optimal

Table 4 Sensitivity of the expected total cost to the variations in the cancellation and postponement costs

Cancellation Cost	Postponement Cost				
	600	1200	1800	2400	3000
600	41063	65547	88579	110614	131229
1200	41135	66012	88464	110963	131908
1800	41141	66053	89340	111244	132185
2400	41143	66069	89375	111471	132453
3000	41142	66080	89382	111491	132661

Table 5 Sensitivity of the expected number of cancellations to the variations in the cancellation and postponement costs

Cancellation Cost	Postponement Cost				
	600	1200	1800	2400	3000
600	15.90	26.90	24	21.75	21.75
1200	0.40	9.05	19.20	20.95	16.10
1800	0.20	0.50	8.55	18.10	15.25
2400	0.00	0.40	0.70	8.55	9.60
3000	0.00	0	0.40	0.40	3.90

solution postpones surgery to later days in the time window to prevent future cancellations from occurring. In other words, low postponement cost provides greater flexibility in scheduling. Since the low postponement cost leads to a lower number of cancellations, the services performing surgeries of lower urgency are likely to observe fewer cancellations.

The second factor is important when both the costs of cancellation and postponement are high. In such cases, the optimal solution schedules surgeries to a day as early as possible, so that a high postponement cost is avoided. However, the selected day should also minimize the chance of cancellation because of the high cancellation cost. If the overtime cost is not excessive, the number of cancellations can still be reduced even when the postponement cost is high.

Tables 6 and 7 illustrate the sensitivity of the expected total cost and average number of cancellations to the changes in the cancellation and overtime costs (postponement

Table 6 Expected total cost as a function of the cost of cancellation and overtime costs (under constant postponement cost)

Postponement cost=1800	Overtime Cost		
Cancellation cost	30	60	90
600	76098	88579	98076
1200	76370	88664	98584
1800	76610	89340	99036
2400	76622	89375	99077
3000	76622	89382	99098

Table 7 Expected number of cancellations as a function of the cost of cancellation and overtime costs (under constant postponement cost)

Postponement cost=1800	Overtime Cost		
Cancellation cost	30	60	90
600	15.74	24	31.6
1200	9.6	19.2	26.1
1800	4.5	8.55	9.35
2400	0	0.6	0.6
3000	0	0.4	0.4

cost is constant in these cases). The expected total cost increases significantly when only the overtime cost increases, but it increases only slightly when only the cancellation cost increases. It is also evident that the number of cancellations increases as the overtime cost increases, especially when the postponement cost is not lower than the cancellation cost. Table 7 also suggests that the first factor, ratio of postponement to cancellation cost, has a significant influence on the number of cancellations.

5.3.2. Sensitivity to variability in demand. In this section, we evaluate the impact of uncertainty on the expected total cost and the number of surgery cancellations. Table 8 shows a comparison among six different moderate size instance sets. Note that the expected cumulative duration of the requested surgeries is held nearly constant (i.e. 9 hours), while the standard deviation of the durations is varied over different sets. The values for the cost parameters are fixed for all instance sets ($c^o = 60, c^i = 600, l_i = 1200$).

Table 8 shows that the expected total cost increases as the variation in the system increases. The table also indicates that when there is no variation, cancellations are not observed. When the coefficient of variation is increased from 0 to 0.07 by increasing the standard deviation from $\sigma = 0$ to $\sigma = 39$ cancellations are still not observed. However, as the standard deviation further increases, the number of cancellations also increases significantly. The results suggest that removing the variation in requests over a planning period will reduce the number of cancellations. It also suggests that in the surgical services whose surgery mixes may change significantly from day to day, cancellations are more likely to occur. Therefore, grouping similar types of surgeries and assigning them to the same day and OR may reduce cancellations.

Table 8 Variation in the expected number of cancellations and total cost due to change in the standard deviation (σ^{ins}) of daily cumulative surgery durations (with a mean of 540 minutes for all instance sets) over the planning period

	Standard Deviation (σ^{ins})					
	0	39	78	117	156	234
Expected cancellations	0	0	3.5	8.79	15.7	26.9
Expected total cost	51080	54126	54357	54873	58043	65547

5.4. Value of the Stochastic Solution

We compared the solutions found for the stochastic model with the solutions of a deterministic heuristic to estimate the value of the stochastic solution (VSS). The heuristic is an extension of the first fit decreasing heuristic, which is a well known heuristic for bin packing problems. The heuristic is myopic in the sense that it does not consider future outcomes while giving decision on the assignment of surgeries into future.

At each stage, the surgeries requested at that stage are ordered from longest to shortest expected duration. Next, the surgeries are assigned to a future day and OR, consecutively, according to the order in the surgery list. The heuristic attempts to assign a surgery to the earliest day available within the allowable time window. The availability of the day depends on the remaining capacities of the ORs which are appropriate for assignment in terms of equipment restrictions. A capacity threshold is set for the ORs to prevent having high overtime. The thresholds are defined such that at most one surgery can be performed in an OR during the overtime period. The heuristic attempts to assign a surgery to the OR opened earliest. If a surgery can not be assigned to any open OR on a particular day, then a new OR is opened to assign the surgery. If there is no additional OR available to open on the same day, then the next day is considered. If the next day is outside of the allowable time window, the surgery is performed during the overtime hours in one of the day and OR combinations. The steps of the heuristic are summarized as follows:

Bin Packing Heuristic

for $t = 1$ to H

Sort the surgeries requested at stage t from longest to shortest duration to form the sorted list, L. Let L_i be the surgery in the i^{th} order, and n^L be the size of the ordered list

```
  for i = 1 to n^L
    while surgery assigned = false
      for u = (t + g_{L_i}) to (t + g_{L_i} + h_{L_i})
        for j = 1 to O
          if Equipment constraint is not violated
            if Capacity constraint is not violated
              surgery assigned=true
            end if
          end if
          if surgery assigned = false
            if There is no more additional OR to open
              surgery assigned = true
            end if
          end if
        end for
      end for
    end while
  end for
end for
```

Table 9 compares the EPHA and the heuristic according to the solution quality and computation time based on 20 instances of a large test case. From Table 9, we conclude that the EPHA improves the quality of solutions found by the heuristic significantly. Note that the average and maximum gap between the objective values for the EPHA and heuristic solutions are 13% and 16%, respectively. Note that the heuristic is very fast with a CPU time of less than 1 second. On the other hand, on average, 3641 CPU seconds are needed to obtain a solution using the EPHA.

Table 9 Comparison of the EPHA with the first fit decreasing heuristic

	Expected Cost (in dollars)		CPU time (in seconds)	
Instance #	PHA	Heuristic	PHA	Heuristic
1	119640	139112	10800.00	0.00
2	134835	156825	455.20	0.00
3	100200	113034	215.52	0.00
4	104895	118103	1049.64	0.00
5	92608	110138	732.46	0.00
6	84840	94073	280.94	0.00
7	148298	171759	10800.00	0.00
8	131318	147066	833.50	0.00
9	77880	87323	383.44	0.00
10	139455	160043	10800.00	0.00
11	94883	108079	10800.00	0.00
12	116811	130449	1198.13	0.00
13	124733	148950	335.01	0.00
14	123953	142386	10800.00	0.00
15	94965	107764	248.73	0.00
16	105578	121044	598.52	0.00
17	129518	142661	10800.00	0.00
18	108795	127733	468.78	0.00
19	136365	154643	391.47	0.00
20	116096	132773	821.78	0.00

6. Conclusions

In this article, we proposed a multi-stage stochastic mixed integer programming formulation for the allocation of surgeries to ORs under uncertainty over a finite planning horizon. We first implemented an extension of PHA, called EPHA, and compared EPHA solutions with the solutions found using CPLEX 12. We then analyzed the trade-offs between cancellation, postponement, and overtime costs with respect to their impact on total expected costs and surgery cancellations. We also assessed the impact of varying levels of uncertainty. Finally, we compared an easy to implement heuristic with the EPHA to estimate the benefit of considering uncertainty in the surgery planning and scheduling process. Following are the most significant findings of our study:

- The ratio of postponement cost to cancellation cost is one of the factors that significantly impacts the expected number of cancellations. If the postponement cost is higher than the cancellation cost, then it is likely to observe surgery cancellations. A relatively low postponement cost allows a wider time window to consider for scheduling surgeries, thus helping to reduce the cancellations. Cancellations are also higher for the cases where overtime is favored due to high levels of postponement and cancellation costs. Another factor to which the expected number of cancellations is highly sensitive is the level of uncertainty in demand and total daily surgery duration. Cancellations do not exist under a deterministic case, but increase as the variation in demand increases.
- Due to the significant effect of uncertainty, the EPHA outperforms a deterministic heuristic that maybe used in practice. The results suggest that a good heuristic must carefully handle the variation in surgery requests, because the cancellations can be reduced when surgeries with similar durations are performed together. Thus, there may be benefits to studying heuristics which minimize the maximum variance of total surgery durations over ORs and days.
- The initial values of Lagrangian multipliers, Lagrangian multiplier and penalty update methods are the factors that significantly affect the performance and solution quality of the PHA. Our Lagrangian multiplier update method maintained the convergence of the EPHA. Our penalty update method accelerated convergence, but had a slight negative effect into the solution quality. The EPHA requires a short time to yield near optimal solutions for difficult problem instances that can not be solved by CPLEX within a reasonable amount of time.

Our model can be used in practice for an extended length of planning periods (i.e $P \geq H$) using a rolling horizon approach. One can solve the surgery planning model every H days for a planning period of P days. The requested surgeries that are not performed on any day in an H-day period may be reconsidered for scheduling along with the new surgery requests of the next H-day period. The urgency level of the leftover surgeries can be increased by changing the cancellation and postponement costs of them before solving the model for the new period. This approach will be the next step of our future studies.

Acknowledgments

This material is based in part upon work supported by the National Science Foundation under Grant Number CMMI 0844511. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Argo, J. L., C. C. Vick, L.A. Graham, K.M.F. Itani, M.J. Bishop, M.T. Hawn. 2009. Elective surgical case cancellation in the veterans health administration system: identifying areas for improvement. *The American Journal of Surgery* **198** 600–606.

Barro, D., E. Canestrelli. 2005. Dynamic portfolio optimization: Time decomposition using the maximum principle with a scenario approach. *European Journal of Operational Research* **163** 217–229.

Crainic, T.G., X. Fu, M. Gendreau, W. Rei, S.W. Wallace. 2011. Progressive hedging-based metaheuristics for stochastic network design. *Networks* **58** 114–124.

Fei, H., C. Chu, N. Meskens. 2009. Solving a tactical operating room planning problem by a column generation based heuristic procedure with four criteria. *Annals of Operations Research* **166** 91–108.

Fei, H., C. Chu, N. Meskens, A. Artiba. 2008. Solving surgical cases assignment problem by a branch-and-price approach. *International Journal of Production Economics* **112** 96–108.

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

Gerchak, Y., D. Gupta, M. Henig. 1996. Reservation planning for elective surgery under uncertain demand for emergency surgery. *Management Science* **42** 321–334.

Gillen, S. M. I., K. Catchings, L. Edney, R. Prescott, S. M. Andrews. 2009. What's all the fuss about? day-of-surgery cancellations and the role of perianesthesia nurses in prevention. *Journal of Perianesthesia Nursing* **26** 396–398.

Guinet, A., S. Chaabane. 2003. Operating theatre planning. *International Journal of Production Economics* **85** 69–81.

Gul, S., B. Denton, J. Fowler, T. Huschka. 2011. Bi-criteria scheduling of surgical services for an outpatient procedure center. *Production and Operations Management* **20** 406–417.

Hans, E., G. Wullink, M. van Houdenhoven, G. Kazemier. 2008. Robust surgery loading. *European Journal of Operational Research* **185** 1038–1050.

Haugen, K.K., A. Lokketangen, D.L. Woodruff. 2001. Progressive hedging as a meta-heuristic applied to stochastic lot-sizing. *European Journal of Operational Research* **132** 116–122.

Helgason, T., S.W. Wallace. 1991. Approximate scenario solutions in the progressive hedging algorithm. *Annals of Operations Research* **31** 425–444.

HFMA. 2005. Achieving operating room efficiency through process integration. *Health Care Financial Management Association Report* .

Hvattum, L.M., A. Lokketangen. 2009. Using scenario trees and progressive hedging for stochastic inventory routing problems. *Journal of Heuristics* **15** 527–557.

Lamiri, M., F. Grimaud, X. Xie. 2009. Optimization methods for a stochastic surgery planning problem. *International Journal of Production Economics* **120** 400–410.

Lamiri, M., X. Xie, A. Dolgui, F. Grimaud. 2008a. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research* **185** 1026–1037.

Lamiri, M., X. Xie, S. Zhang. 2008b. Column generation approach to operating theater planning with elective and emergency patients. *IIE Transactions* **40** 838–852.

Listes, O., R. Dekker. 2005. A scenario aggregation-based approach for determining a robust airline fleet composition for dynamic capacity allocation. *Transportation Science* **39** 367–382.

Min, D., Y. Yih. 2010. Scheduling elective surgery under uncertainty and downstream capacity constraints. *European Journal of Operational Research* **206** 642–652.

Mulvey, J.M., H. Vladimirou. 1991a. Applying the progressive hedging algorithm to stochastic generalized networks. *Annals of Operations Research* **31** 399–424.

Mulvey, J.M., H. Vladimirou. 1991b. Solving multistage stochastic networks: An application of scenario aggregation. *Networks* **21** 619–643.

Mulvey, J.M., H. Vladimirou. 1992. Stochastic network programming for financial planning problems. *Management Science* **38** 1642–1664.

Rockafellar, R.T. 1976. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization* **14** 877–898.

Rockafellar, R.T., R.J.B. Wets. 1991. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research* **16** 119–147.

Santos, M.L.L.D., E.L.D. Silva, E.C. Finardi, R.E.C Goncalves. 2009. Practical aspects in solving the medium-term operation planning problem of hydrothermal power systems by using the progressive hedging method. *Electrical Power and Energy Systems* **31** 546–552.

Takriti, S., J.R. Birge. 2000. Lagrangian solution techniques and bounds for loosely coupled mixed-integer stochastic programs. *Operations Research* **48** 91–98.

Takriti, S., J.R. Birge, E. Long. 1996. A stochastic model for the unit commitment problem. *IEEE Transactions on Power Systems* **11** 1497–1508.

Wallace, S.W., T. Helgason. 1991. Structural properties of the progressive hedging algorithm. *Annals of Operations Research* **31** 445–456.

Watson, J.P., D.L. Woodruff. 2011. Progressive hedging innovations for a class of stochastic mixed-integer resource allocation problems. *Computational Management Science* **8** 355–370.

Zonderland, M.E., R.J. Boucherie, N. Litvak, C.L.A.M. Vleggeert-Lankamp. 2010. Planning and scheduling of semi-urgent surgeries. *Health Care Management Science* .