Set Partitioning Methods for Scheduling: an Application to Operating Theatres

Elizabeth Rowse, Rhyd Lewis, Paul Harper, Jonathan Thompson
Cardiff University
Cardiff, UK
RowseEL@cardiff.ac.uk, LewisR9@cardiff.ac.uk,
Harper@cardiff.ac.uk, ThompsonJM1@cardiff.ac.uk

ABSTRACT

A major factor contributing to the high number of cancelled operations in hospitals is the unavailability of beds on hospital wards for post-operative recovery. By modelling the impact of the operating theatre timetable, the Master Surgery Schedule (MSS), on the demand for beds and vice versa, an MSS can be produced that results in a reduced number of cancelled operations whilst also levelling the demand for beds throughout the week. In this work, a set partitioning formulation has been developed to assign surgical specialties to operating theatres, and a novel extension of the model has been used to incorporate constraints on the demand for post-operative beds. Initial results are presented to highlight the potential of this model, and the use of a robust optimisation approach is also investigated in order to address the stochastic nature of the factors affecting operating theatre scheduling.

KEYWORDS. Scheduling. Set partitioning. Optimisation. Simulation. Healthcare.

Main Area: Healthcare

1. Introduction

The scheduling of hospital operating theatres (OTs) has proven to be a very challenging and active area of research over the past 20 years. The practical importance of the problem is due to the surgical suite being one of the major departments within a hospital. Surgical suites have high costs associated with their function and surgical patients provide a significant proportion of demand for other hospital departments, both before and after surgery. It is therefore desirable, from the hospital manager's point of view, that the OTs are utilised in the best way in order to make the most of these valuable resources.

OT scheduling is also an interesting and challenging research problem due to the large number of variable factors that can affect operations. According to Van Oostrum et al. [16], the main uncertainties related to the scheduling of operations are the stochastic duration of surgical procedures, personnel availability, the no-show of patients and the occurrence of emergency patients.

Santibanez et al. [15] discuss the potential benefits of a systematic approach to surgery scheduling, including the increased efficiency of the OTs, increased patient throughput, lower wait times for both patients and surgical staff, and increased transparency and fairness in the allocation of time to different surgical specialties.

There are several stages of OT scheduling that are used for different planning horizons. This paper concerns the scheduling of operations at the tactical level which is used for medium-term planning of elective patients. This consists of the construction of a Master Surgery Schedule (MSS) which is a cyclic weekly timetable used by theatre staff. The MSS specifies the surgical specialty that has priority in each OT during each session of the week. A novel extension to the construction of the MSS is considered in this paper which takes into account the effect the MSS then has on the demand for beds on wards for post-operative recovery. By constructing the MSS in this way, demand for beds on wards can be leveled throughout the week resulting in fewer elective operations being cancelled.

A more in-depth review of previous approaches to construction of the MSS is given in Section 2. The new set partitioning model that incorporates bed constraints is then presented in Section 3, followed by a discussion on the methods used to incorporate the bed constraints in the model in Section 4. Results of this new model are discussed in Section 5 and conclusions are given in Section 6.

2. Literature review

As mentioned previously, OT planning and scheduling continues to be a challenging and popular topic for research. The reader is directed to a particularly thorough literature review paper by Cardoen et al. [7] for an overview of the different stages of OT scheduling. Their literature review found that the most common technique used for the construction of the MSS is mixed integer programming (MIP). Here we review a selection of papers as examples of previous approaches to constructing the MSS.

Blake and Carter [6] commented on the scope of OT scheduling research and advised that techniques that integrate the OTs with other hospital departments are urgently required. The majority of papers still only consider an isolated OT [7], however some chose to integrate other resources such as wards and the intensive care unit. A reason why the scope of research projects may have been simplified to exclude other departments could be the resulting increased formulation and computational complexity of the scheduling models.

A small number of papers in the literature consider constructing the MSS for only one surgical team or only one OT. Vissers et al. [18] use MIP to construct an MSS for the cardiothoracic surgery department with a four week cycle time. A number of resources are considered in the model as constraints, specifically intensive care beds and nursing staff. Kuo et al. [10] also use intensive care beds as constraints in their model, and use linear programming to allocate multiple specialties to multiple OTs in order to maximise surgeon revenue. Their results indicate a 15% increase in revenue.

Van Oostrum et al. [16] use an MIP model to construct an MSS that uses a column generation technique to find a solution. The stochastic nature of the duration of surgical procedures is considered and planned slack is built into the timetable in order to account for this. Their MIP model aims to maximise the OT utilisation as well as levelling the subsequent hospital bed requirements.

Belien and Demeulemeester [3] propose a number of MIP and quadratic programming models for constructing the MSS. They evaluate these methods by considering the resulting bed occupancy after surgeries, with the aim of levelling it as much as possible. They build a model that minimises the total expected bed shortage with constraints on the demand for OT blocks for each surgical group, and constraints on the capacity of the number of available OT blocks each day. Belien et al. [4] subsequently discuss a decision support system for the implementation of these models in a large hospital. They find that the different models provide slightly different schedules, and conclude that it is up to the manager to choose the 'best' schedule.

As an alternative to MIP, Vanberkel et al. [17] use a queuing theory approach to build the MSS in such a way that demand on downstream hospital departments is predicted and taken account in the MSS.

3. Set partitioning model

The set partitioning problem (SPP) is a binary integer programming model that determines how items in a set can be partitioned into smaller subsets such that all items in the larger set are contained in exactly one subset. This model has been successfully used for the modeling of scheduling and rostering problems [14], and also vehicle routing problems [2]. In general, the SPP is NP-hard. However, in some cases exact approaches can be used to determine globally optimal solutions [11], even for instances involving hundreds of millions of variables and hundreds of constraints [12]. Structural properties of the constraint matrix can also be taken advantage of in order to solve large SPP models [1].

The traditional formulation of the SPP model is:

$$\begin{aligned} \min\ z &= c^T x && (1) \\ \text{s.t. } Ax &= e && (2) \\ x &\in \{0,1\}^n && (3) \end{aligned}$$

where A is an $m \times n$ binary matrix, c is an arbitrary vector of costs, and $e = (1, 1, 1, ..., 1)^T$ is an m-vector. The decision variables x_j, $j = 1, ..., n$, can be thought of as the probability that the j^{th} column is included in a solution.

There are some models in the literature that deviate from the pure SPP model structure and these generalised set partitioning models are often found in crew rostering applications [12]. In these models the right-hand-side vector, e, need not be a unit vector, and some constraints do not need to be equalities. It has been shown that there are benefits to constraints of this type with relation to finding integer basic feasible solutions in the LP relaxation of the set partitioning model [13].

Set partitioning models have previously been used in healthcare applications to assign elective surgical patients to OT slots taking into account constraints relating to OT and surgeon availability [9], and to schedule anesthesiologists for surgery based on the matching of skills with specific tasks [8]. However, it is not apparent that set partitioning models have been used for a more tactical level of OT scheduling such as the construction of the MSS.

The proposed set partitioning model for the application of the construction of the MSS is outlined herein. The aim of the model is to select a subset of possible 'plans' for each surgical specialty in order to assign one specialty to each OT session and to ensure that the number of predicted patients in post-op recovery does not exceed the number of beds available on wards. A plan for a specialty defines which OT the specialty has use of on which day of the week and during which session (morning/afternoon/whole day), and reflects the specialty's preferences of theatres, days and distribution of sessions. All possible plans are generated using an enumeration algorithm whilst taking into account these preferences. The solution of the model will provide one plan for each of the surgical specialties, which when put together, will form the MSS.

The basic formulation of the proposed set partitioning model for the construction of the MSS is:

$$\begin{aligned} \min\ z &= c^T x && (4)\\ \text{s.t. } Ax &= e && (5)\\ Bx &\leq d && (6)\\ x &\in \{0,1\}^n && (7) \end{aligned}$$

Where x_j, $j = 1, ..., n$, are decision variables that indicate whether or not a plan is selected in the final solution and c is a vector of the cost of each plan.

The 'cost' of each plan reflects the difference between the total number of beds available for post operative care and the predicted bed occupancy. The objective is then to minimise this cost, i.e. it is desirable to minimise the number of empty beds on the wards.

A is an $m \times n$ matrix of zeros and ones where the columns represent possible plans for each surgical specialty. The first s rows of A represent generalised upper bound (GUB) constraints that relate each plan to a specific specialty. The remaining rows represent constraints for each OT session. These rows of the A matrix have elements

$$a_{ij} = \begin{cases} 1 & \text{if operating room session } i \text{ is used in plan } j \\ 0 & \text{otherwise} \end{cases}$$

The right-hand side values of the OT constraints are given in the vector e where $e = (1, 1, 1, ..., 1)^T$. This indicates that only one plan must be selected in the solution for each specialty, and that only one specialty can occupy an operating room session at any one time. The matrix B has entries that are determined from the plans in the A matrix and represent the number of surgical patients who require beds for each plan on each ward on each day. The elements of the B matrix, $b_{kj}^{(l)}$, represent the number of beds required on ward k on day l for plan j. If there are p wards and the model considers bed usage on q days, then B is a $(p \times q) \times n$ matrix. Constraints are constructed so that the number of beds required is less than or equal to the number of beds available. The number of beds available on ward k on day l, $d_k^{(l)}$, form the right-hand side values of the bed constraints.

The bed constraints are treated as elastic in order to allow for the 'sharing' of beds between specialties on different wards. Both slack and surplus variables are added to the constraints so that they can be under or over satisfied. In this particular application, slack variables represent empty beds, and surplus variables represent additionally required beds. In order to be able to specify how many beds are transferred between which particular wards, consider a square $p \times p$ matrix, $Z^{(l)}$, whose elements $z_{kv}^{(l)}$ are decision variables that specify how many beds are moved from ward k to ward v on day l.

By definition, it follows that the sum of the elements of row k of $Z^{(l)}$ represents the number of empty beds on ward k on day l. Similarly, following from the definition of the surplus variables for this application, the sum of the elements of column k of $Z^{(l)}$ represents the number of extra beds for ward k on day l.

In order to control which slack and/or surplus variables are used in the bed constraints, a matrix is used to define allowable movements of patients between wards. This transition matrix, W, is a square $p \times p$ matrix that is not necessarily symmetric. W is constructed based on prior knowledge obtained from the hospital on which wards each specialty can use, and so W is assumed to be constant for each day l. The elements of W are:

$$w_{kv} = \begin{cases} 1 & \text{if patients meant to be on ward } k \text{ are able to use beds on ward } v \\ 0 & \text{otherwise} \end{cases}$$

Combining elements in W and $Z^{(l)}$ gives the total slack and surplus for each ward, and results in constraint (3) being re-written as follows:

$$\sum_{j=1}^{n} b_{kj}^{(l)} x_j - \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} + \sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} = d_k^{(l)} \quad \forall \quad k = 1, ..., p, \;\; l = 1, ..., q \tag{8}$$

An additional constraint is that the sum of the surplus variables across all wards does not exceed the sum of the slack variables across all wards on each day. This is needed in the model to prevent the total number of beds in the hospital from being exceeded.

The formulation of the set partitioning model for the construction of the MSS is therefore:

$$\min \quad \sum_{j=1}^{n} c_j x_j \tag{9}$$

$$\text{s.t.} \quad \sum_{j=1}^{n} a_{ij} x_j = 1 \quad \forall \quad i = 1, ..., s \tag{10}$$

$$\sum_{j=1}^{n} a_{ij} x_j \leq 1 \quad \forall \quad i = s+1, ..., m \tag{11}$$

$$\sum_{j=1}^{n} b_{kj}^{(l)} x_j - \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} + \sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} = d_k^{(l)} \quad \forall \quad k = 1, ..., p, \quad l = 1, ..., q \tag{12}$$

$$\sum_{k=1}^{p} \sum_{v=1}^{p} w_{kv} z_{vk}^{(l)} \leq \sum_{k=1}^{p} \sum_{v=1}^{p} w_{vk} z_{kv}^{(l)} \quad \forall \quad l = 1, ..., q \tag{13}$$

$$x_j \in \{0, 1\} \quad \forall \quad j = 1, ..., n \tag{14}$$

$$z_{kv}^{(l)} \geq 0 \text{ and integer} \quad \forall \quad k = 1, ..., p, \quad v = 1, ..., p, \quad l = 1, ..., q \tag{15}$$

4. Generating the B matrix

As mentioned, the A matrix defines in which theatre and at what time each surgical specialty will operate. Using the A matrix, the B matrix is then generated by filling in the required number of beds on each ward on each day for each plan. In the model, patients are sent from theatre to either a ward or the critical care unit for post-operative recovery. Patients who are sent to critical care are subsequently moved to a ward. Data concerning patients post-operative length of stay (LoS) in each ward and critical care for all specialties can be used to determine how long patients will remain in beds in the model.

In our case, two different methods, based on the use of the LoS data are used to populate the B matrix. These methods are:

1. Single scenario of sampled LoS
2. Multiple scenarios of sampled LoS

Let the post-op length of stay for a patient be denoted by the random variable T. More specifically, T denotes the duration of time after surgery until the patient either leaves hospital (end of the spell in hospital) or moves to the care of a different specialty (end of episode). In either case, T is effectively the time taken for the patient to 'recover' from surgery.

In the first method of filling in the B matrix, an 'example' of the number of beds required is generated based on each plan. The conditional probability of failure is used from survival analysis to determine the probability that a patient leaves on day d given that the post-op LoS has reached d days. The Kaplan-Meier estimate of the survivor function is used because no parametric distribution can be fit to our LoS data. Using the random variable T as the post-op LoS, the survivor function, denoted $S(t)$, is the probability that the random variable T takes a value longer than a specified time t, i.e. $S(t) = P(T > t)$. An estimate of $S(t)$ can be calculated using the Kaplan-Meier method, sometimes known as the product-limit estimate. The Kaplan-Meier estimate of the survivor function, $S(t)$, at time $t_{(j)}$, is given by $\hat{S}(t_{(j)}) = \hat{S}(t_{(j-1)}) \times P(T > t_{(j)} \mid T \geq t_{(j)})$, where $t_{(j)}$ are the ordered survival times/length of stays for $j = 0, 1, 2, ..., k$. Using information from the LoS data such as the number of patients that have each distinct length of stay, $m_{(j)}$, and the number of patients that could have left at each length of stay time, $n_{(j)}$, the Kaplan-Meier estimate is calculated as:

$$\hat{S}(t_{(j)}) = \frac{n_{(j)} - m_{(j)}}{n_{(j)}}$$

The conditional probability of failure, $L(t_{(j)})$, is the probability that the event (patient leaves hospital) occurs in a small time interval h after time t, and is defined as $L(t_{(j)}) = P(t < T < t + h \mid T > t)$. $L(t_{(j)})$ can be estimated when finding the Kaplan-Meier estimate of $S(t)$ as follows:

$$\hat{L}(t_{(j)}) = \frac{m_{(j)}}{n_{(j)}}$$

The B matrix is then generated using the conditional probability of failure estimate, $\hat{L}(t_{(j)})$, according to the following algorithm.

Algorithm 1 Generate B Matrix for Single Scenario

```
for each plan (column) in A matrix do
  Look-up which specialty the plan refers to
  for each session (row) of A matrix do
    if the specialty is scheduled in the OR session then
      Enter the number of new arrivals in the row in the B matrix that corresponds to
      the weekday that the OR session is on
      for each day in the length of stay distribution for this specialty do
        for each remaining arrival do
          Generate a random number, r ∈ [0, 1]
          if r > P(leaving hospital on this day, given that the patient has been in
          hospital this long) then
            Decrease the number of remaining arrivals by 1
          end if
        end for
        Update B matrix with number of remaining arrivals on this day
      end for
    end if
  end for
end for
```

The second method of populating the B matrix is very similar to the first, however instead of just one scenario, there are multiple scenarios, say t, scenarios. This method essentially generates t different B matrices using the same technique as discussed for method 1, and appends them to create more bed constraints for the optimisation model. The idea behind using multiple scenarios is that the more constraints that an optimal schedule can satisfy, the more likely the schedule will be able to cope with different realisations of uncertain bed demands when implemented in a hospital. The number of scenarios, t, is chosen by the user, however t must be chosen with care as there is a trade-off between including more scenarios to result in a more 'robust' schedule, and having too many constraints so that feasible solutions cannot be found. If t scenarios are generated, then there becomes

$t \times p \times q$ bed constraints, where p wards and q days are being modelled. Constraints (13) and (14) are re-formulated as below to reflect the multiple scenarios in this method for all scenarios $g = 1, ..., t$, wards $k = 1, ..., p$ and days $l = 1, ..., q$.

$$\sum_{j=1}^{n} b_{gkj}^{(l)} x_j - \sum_{v=1}^{p} w_{kv} z_{gvk}^{(l)} + \sum_{v=1}^{p} w_{vk} z_{gkv}^{(l)} = d_{gk}^{(l)} \tag{16}$$

$$\sum_{k=1}^{p} \sum_{v=1}^{p} w_{kv} z_{gvk}^{(l)} \leq \sum_{k=1}^{p} \sum_{v=1}^{p} w_{vk} z_{gkv}^{(l)} \tag{17}$$

5. Results

A large teaching hospital in Wales, UK, in which over 25,000 surgical operations are performed annually, is used as a case study. Around 18% of operations in this hospital are cancelled annually, with non-clinical reasons, such as lack of available beds, accounting for 54% of these. In addition, 31% of patients are currently placed on a ward that does not necessarily have specialist nurses or equipment for the surgical specialty post-surgery due to shortfalls in bed capacity on the specialist ward.

The surgical suite at the case study hospital has 14 OTs used by 18 surgical specialties. Elective operations are scheduled during two operating sessions per day (morning/afternoon) over a five day week (Monday - Friday). There is one dedicated emergency theatre in which no elective operations are scheduled. Data on the preferences on theatres and sessions has been obtained from the hospital managers and has been used to generate the A matrix. The enumeration of all possible plans for this case study results in an A matrix with almost 1500 plans (and hence 1500 binary decision variables) and 158 OT constraints. Ten wards and a critical care unit are modelled for seven days a week, resulting in a B matrix with 77 bed constraints for a single scenario (method 1).

The optimisation model was run and simulations of the resulting MSS was then performed in order to obtain a measure of its robustness. A simulation of the optimal schedule is intended as a test to determine how the schedule will cope when different aspects of the uncertainty are realised i.e. how many patients will require a bed on each day in each ward compared to the number of beds available. If more patients require a bed than there are beds available it is equivalent to a violated bed constraint occurring in the optimisation model. It is of interest to examine the percentage of simulations of the optimal schedule for which at least one of the bed constraints would be violated. Results from 1000 runs are presented for up to 10 scenarios in Figure 1 below.

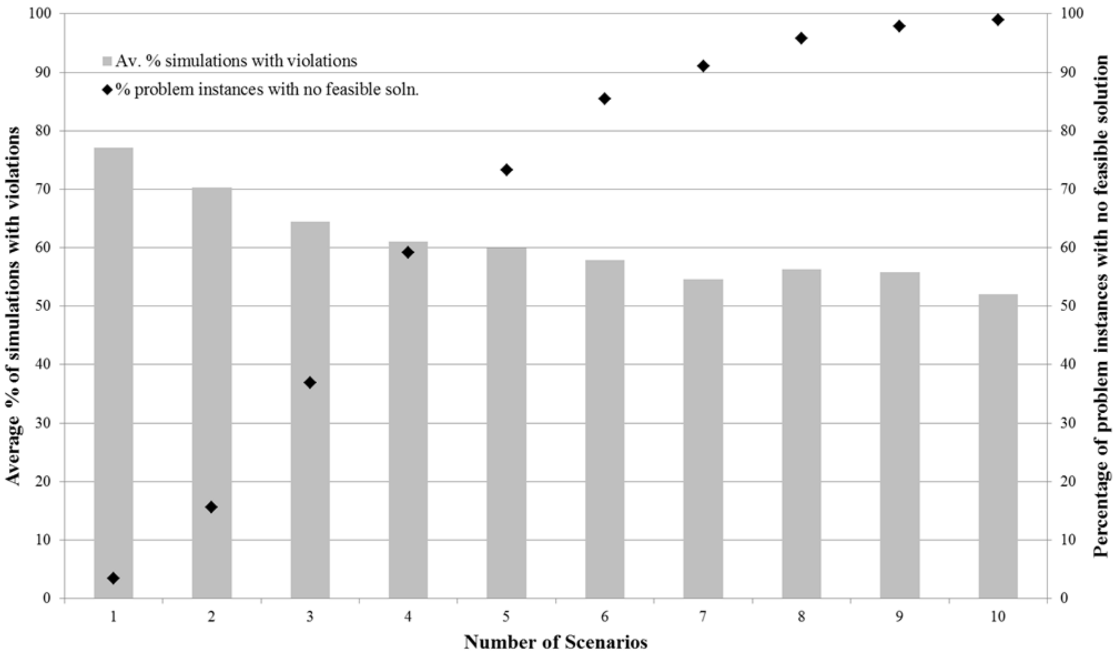

Figure 1 - Graph to show the results of 1000 runs

As can be seen from Figure 1, the average percentage of simulations, in which the optimal schedule would have resulted in at least one of the bed constraints being violated, decreases as the number of scenarios increases. This is as expected, due to more bed constraints being included in the optimisation acting as realisations of uncertainty; the resilience of the optimal schedule to uncertainty in the simulations will increase with the number of bed constraints included.

It can also be seen that as the number of scenarios increases, and hence the number of bed constraints in the model increases, the percentage of problem instances for which no feasible solution exists increases. This is also as expected due to the problems becoming more constrained as scenarios are added to the model.

6. Conclusions

A set partitioning based optimisation model has been developed that includes a novel extension to the formulation to incorporate constraints on the demand for beds. The model constructs an MSS by assigning specialties to OTs subject to post-operative bed constraints. An increasing number of scenarios have been used to generate bed constraints for the optimisation model; simulation of the resulting optimal MSSs has then been performed in order to obtain a measure of the robustness of the MSS. The initial results presented here indicate that the more scenarios that are included in the bed constraints, the more robust the resulting MSS is with respect to the percentage of simulations that have violated bed constraints. The more scenarios that are included, the lower the percentage of simulations that have violated bed constraints. However, there is a trade-off between the number of scenarios to include and the likelihood of finding a feasible solution, since the number of problem instances for which no feasible solution exists increases as the number of scenarios increases.

The method of filling in the B matrix using multiple scenarios of sampled LoS can be used to achieve a more robust scheduling approach. A robust schedule is desirable from the hospital's point of view because it will give more confidence that the schedule can withstand the variations in demand for beds when different aspects of this uncertainty are realised. Current work is focusing on adapting the SPP model presented here into a more robust formulation in the sense of Bertsimas and Sim [5].

Acknowledgments

The first author would like to thank Professor David Ryan and other members of the Department of Engineering Science, University of Auckland, namely Associate Professor Andrew Mason and Dr Michael O'Sullivan, for their support and advice on this work. Thanks also go to Tony Turley and Andrew Nelson at the case study hospital for providing information and data related to this problem.

References

[1] E. Balas and M. W. Padberg. Set partitioning: A survery. *SIAM Review*, 18(4):710–760, 1976.

[2] M. L. Balinski and R. E. Quant. On an integer program for a delivery problem. *Operations Research*, 12(2):300–304, 1964.

[3] J. Belien and E. Demeulemeester. Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research*, 176(2):1185–1204, 2007.

[4] J. Belien, E. Demeulemeester, and B. Cardoen. A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling*, 12(2):147–161, 2009.

[5] D. Bertsimas and M. Sim. The price of robustness. *Operations Research*, 52(1):35–53, 2004.

[6] J. T. Blake and M. W. Carter. Surgical process scheduling: a structured review. *Journal of the Society for Health Systems*, 5(3):17–30, 1997.

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

[8] E. A. Ernst, L. S. Lasdon, L. E. Ostrander, and S. S. Divell. Anesthesiologist scheduling using a set partitioning algorithm. *Computers and Biomedical Research*, 6(6):561–569, 1973.

[9] H. Fei, C. Chu, and N. Meskens. Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Annals of Operational Research*, 166(1):91–108, 2009.

[10] P. C. Kuo, R. A. Schroeder, S. Mahaffey, and R. R. Bollinger. Optimization of operating room allocation using linear programming techniques. *Journal of the American College of Surgeons*, 197(6):889–895, 2003.

[11] R. E. Marsten and F. Shepardson. Exact solution of crew scheduling problems using the set partitioning model: Recent successful applications. *Networks*, 11(2):165–177, 1981.

[12] D. Ryan. Solution of massive generalized set partitioning problems. *Journal of the Operational Research Society*, 43(5):459–467, 1992.

[13] D. Ryan and J. C. Falkner. On the integer properties of scheduling set partitioning models. *European Journal of Operational Research*, 35(3):442–456, 1988.

[14] D. Ryan and B. A. Foster. An integer programming approach to scheduling. *Computer Scheduling of Public Transport: Urban Passenger Vehicle and Crew Scheduling, A. Wren (ed.), North Holland, Amsterdam*, pages 269–280, 1981.

[15] P. Santibañez, M. Begen, and D. Atkins. Surgical block scheduling in a system of hospitals: an application to resource and wait list management in a british columbia health authority. *Health Care Management Science*, 10(3):269–282, 2007.

[16] J. M. van Oostrum, M. van Houdenhoven, J. L. Hurink, E. Hans, G.Wullink, and G. Kazemier. A master surgical scheduling approach for cyclic scheduling in operating room departments. *OR Spectrum*, 30(2):355–374, 2008.

[17] P. T. Vanberkel, R. J. Boucherie, E. W. Hans, J. L. Hurink, W. A. M. van Lent, and W. H. van Harten. An exact approach for relating recovering surgical patient workload to the master surgical schedule. *Journal of the Operational Research Society*, 62(10):1851–1860, 2011.

[18] J. M. H. Vissers, I. J. B. F. Adan, and J. A. Bekkers. Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management Mathematics*, 16(3):281–304, 2005.