Scheduling the hospital-wide flow of elective patients

Daniel Gartner and Rainer Kolisch

TUM School of Management
Technische Universität München, Germany
e-mail: daniel.gartner@wi.tum.de, rainer.kolisch@wi.tum.de

Abstract

In this paper, we address the problem of planning the patient flow in hospitals subject to scarce medical resources with the objective of maximizing the contribution margin. We assume that we can classify a large enough percentage of elective patients according to their diagnosis-related group (DRG) and clinical pathway. The clinical pathway defines the procedures (such as different types of diagnostic activities and surgery) as well as the sequence in which they have to be applied to the patient. The decision is then on which day each procedure of each patient's clinical pathway should be done, taking into account the sequence of procedures as well as scarce clinical resources, such that the contribution margin of all patients is maximized. We develop two mixed-integer programs (MIP) for this problem which are embedded in a static and a rolling horizon planning approach. Computational results on real-world data show that employing the MIPs leads to a significant improvement of the contribution margin compared to the contribution margin obtained by employing the planning approach currently practiced. Furthermore, we show that the time between admission and surgery is significantly reduced by applying our models.

Keywords: Clinical pathways, Diagnosis-related groups, Patient flow management

1 Introduction

For many years, cost reimbursement has been the standard payment scheme for hospitals. In this scheme, a hospital receives the total cost for treating the patient which is calculated by multiplying the patient's length of stay with a (ward specific) daily rate and adding the costs for clinical procedures applied, such as diagnosis and surgery. Cost reimbursement does not provide an incentive for hospitals to operate efficiently. As a consequence, hospital costs as the largest part of total health care costs (approximately 31% in the U.S., see Lim et al. [31]), have increased sharply. In an effort to limit hospital costs and to create an incentive for hospitals to operate more efficiently, many countries have introduced payment schemes that are based on diagnosis-related groups (DRGs). In these schemes, patients are classified into DRGs with homogeneous clinical characteristics and resources required during treatment within each group, while between groups, the patients' clinical characteristics and therefore costs are different. Hospitals receive payments based on the DRG instead of the applied procedures and the length of stay. The reimbursement the hospital receives for treating a patient with a specific DRG equals the average cost which accrued in a representative sample of hospitals in the year before last. For each country, there is an institution, such as the German institute for the reimbursement in hospitals INEK (see Schreyögg et al. [48]), responsible for generating the sample, collecting, processing and distributing the cost data. With DRG-based payment schemes in place, in order to be profitable a hospital has to operate such that its patient specific costs are lower than the average costs of the hospitals in the sample. However, the cost structure in hospitals, as in many service industries, is such that the predominant share of the operational costs is fixed. Hence, costs can only be marginally reduced by operational decisions. Instead, hospitals have to use existing resources such that revenue or the total contribution margin is maximized. This approach is well-known in the service industry (see Kimes [27] for an example in the service industry in general as well as Schütz and Kolisch [49] for an example in the health care industry in particular).

In this paper, we propose two models to plan the patient flow in hospitals subject to scarce medical resources with the objective to maximize the contribution margin. We assume that we can classify a large enough percentage of elective patients according to DRG and clinical pathway. The clinical pathway defines the procedures (such as different types of diagnostic activities and surgery) as well as the sequence in which they have to be applied to the patient. The decision is then on which day each procedure of each patient's clinical pathway should be done, taking into account the sequence of procedures as well as scarce clinical resources, such that the contribution margin is maximized. We consider a particular variant of the DRG-based payment scheme which is in use in Australia (see Duckett [17]), Switzerland (see Zaugg et al. [58]), France and Germany (see Schreyögg et al. [48]). In this variant, payment

is not only dependent on DRG but also, to some extent, on length of stay. More precisely, the fixed reimbursement is reduced by some fixed amount per day if the length of stay (LOS) falls below a lower threshold value, referred to in the literature as low LOS trim point. In contrast, the fixed reimbursement is increased by some fixed amount per day if the medically necessary length of stay exceeds a high threshold value, denoted by high LOS trim point. By this, the payment scheme takes outliers into account (see, for example, Schreyögg et al. [48]). The concept of trim points is described in Section 3; for more information about DRG trim points, see Busse et al. [5]. An example of a cost, revenue and contribution margin function is given in Section 3.1, while a detailed example of two patients and their corresponding contribution margin for two specific DRGs is provided in Section 3.3. The payment scheme with fixed payment not coupled to the length of stay is a special case of the one that is treated in our study. Hence, our models can be applied to both variants.

Within the health care planning matrix of Hans et al. [23] our planning approach can be classified as offline operational resource capacity planning. It comprises the detailed coordination of the activities regarding current (elective) demand (see Hulshof et al. [24]). Our models thus use planning decisions which have been made at higher planning levels such as strategic decisions on case mix, type and capacity of clinical resources as well as tactical decisions on the master surgery schedule. Then, our models generate decisions on the admission day of patients, the day-based assignment of clinical activities to hospital resources and the discharge day of patients. These results are in turn input for decisions to be made on subordinate planning levels such as bed assignment in wards, sequencing of diagnostic activities and surgical cases within a day. Contrary to most planning approaches for the offline operational level proposed so far, our model considers not only one resource such as a diagnostic device or the operating theater, but all scarce clinical resources. Correspondingly, we do not consider a single type of activity within the clinical pathway, such as the surgery, but all activities. We manage to do so by applying the right amount of aggregation in our models with a period length of one day. Since the capacity demand of a single activity in a clinical pathway is usually significantly smaller than a day, we can use expected values instead of distribution functions for most data. However, for some data such as recovery times of patients with a length of multiple days and arrival times of patients we explicitly take the stochastic nature into account by embedding our models in a rolling horizon approach with Bayesian recovery time prediction. Our computational results using real-world data demonstrate that our approaches lead to an improvement of approximately 5% in the contribution margin when compared to the planning approach currently employed in the hospital from which we obtained our data. Also, we show that the time span between admission and surgery is reduced by about 1.5 days and that the length of stay is reduced by about 2 days.

The remainder of this paper is structured as follows. In Section 2, we provide a survey of the

relevant literature and set apart our approach. Section 3 presents the mathematical models and illustrates them by means of examples. In Section 4, we test our models on real-world data from a midsize hospital. First, we provide results for the two static deterministic models of Section 3. Then, we show how the models are embedded within a rolling horizon approach using Bayesian recovery time prediction in order to take into account uncertain information. We provide a computational comparison of the static deterministic and the dynamic stochastic planning, benchmarking both against the solution approach currently employed by the hospital. Finally, we analyze the impact of our approaches on some key figures such as time span between admission and surgery as well as length of stay. Section 5 closes the paper with conclusions.

2   Literature review

There is a rich literature on offline operational capacity planning in hospitals. Table 1 provides an overview of offline operational capacity planning problems for surgical, inpatient and residential care services, employing the taxonomy of Hulshof et al. [24]. We selected articles based on that taxonomy and restricted our search to the application of mathematical programming as a solution technique. Moreover, we excluded staff-to-shift assignment papers and those that were published before 2000. We have classified each article with respect to the use of scarce clinical resources (single vs. multiple) and the consideration of risk (deterministic vs. stochastic).

Single resources	
Operating room (OR)	[2, 10, 14–16, 18, 19, 22, 25, 28–30, 32, 33, 38, 42, 44]
Beds	[9, 13, 46]
Therapy	[45]
Multiple resources	
OR and ward/ICU beds	[1, 6, 7, 11, 20, 36, 39, 52]
Diagnostic and beds	[12]
Diagnostic and therapy	[26]

(a)

Deterministic	[1, 6, 7, 10, 12, 13, 18–20, 22, 25, 26, 33, 36, 39, 42, 44, 45, 52, 56]
Stochastic	
Arrival	[11, 28–30, 38]
Surgery durations	[2, 14–16, 28, 30, 32, 38]
Ward recovery	[9, 11, 46]
Clinical pathway	[11, 57]

(b)

Table 1: Classification of articles by resources (a) and consideration of risk (b)

The table reveals that the majority of the papers consider a single resource such as the operating theater (for a survey see Cardoen et al. [8] and Guerriero and Guido [21]) or wards (Demeester et al. [13]). The planning of diagnostic services is addressed, for example, by Patrick et al. [37] who, in contrast to our work, employ Markov decision processes as a solution approach. Since our paper addresses all scarce hospital resources instead of a single one, we do

not present these approaches in detail. A smaller number of papers take into account multiple resources, see Vanberkel et al. [55] for an overview. The majority of the papers which consider multiple resources deal with the operating room in combination with its upstream or downstream resources. Pham and Klinkert [39] propose a mixed-integer program (MIP) for surgical case scheduling considering scarce resources at the preoperative, perioperative and postoperative stages, solved with the standard solver CPLEX. Augusto et al. [1] consider the problem of scheduling a fixed number of elective cases subject to scarce resources in the operating theater, the post-anesthesia care unit and the patient transportation unit. The problem is modeled as a MIP and solved by using Lagrangian relaxation. Besides the scheduling decision, the model is used to explore the benefit of letting patients recover in the operating room when there is no capacity in the post-anesthesia care unit. Conforti et al. [12] address the so-called "week hospital problem" in which a decision is made if and when elective patients on a waiting list are admitted to the hospital and when the clinical activities of the admitted patients are performed. By definition, the week hospital problem ensures that all admitted patients are discharged in the week they have been admitted. The objective is to maximize the sum of the scores of admitted patients. The planning horizon of one week is divided into periods of a half day length. The paper proposes a MIP which is solved with the standard solver CPLEX.

On a tactical planning level, Min and Yih [34] address the surgery scheduling problem for elective patients taking into account stochastic surgery durations and stochastic capacity of the surgical intensive care unit by employing a stochastic discrete program which is solved with the sample average approximation method. Vissers et al. [57] consider the tactical patient mix optimization problem for a single specialty. Employing a MIP, for each day of the week they decide on the number of patients from different categories to be admitted into the hospital. The scarce clinical resources beds (before surgery and for recovery), the operating theater and the intensive care unit are taken into account. The objective is the minimization of the over- and underutilization of the clinical resources. Apart from the recent approaches in operational capacity planning mentioned above, Sepulveda [50] can be regarded as an early work in evaluating DRG-based cost saving policies using simulation.

The models proposed in this paper can be categorized into and differentiated from the literature on offline capacity planning of multiple clinical resources as follows. First, an aggregated approach is followed in which clinical activities are assigned to big bucket periods with a length of one day. This is similar to Vissers et al. [57] and Conforti et al. [12] and differs from the detailed scheduling approaches of Min and Yih [34] and Pham and Klinkert [39]. However, Vissers et al. [57] decide on the admission date of patients only and assume for the latter a fixed schedule. On the contrary, in this paper patients have to be admitted and their clinical activities are scheduled. Thus, from a hierarchical planning perspective the models developed

in this paper are located on a decision level below Vissers et al. [57] but located above the detailed resource planning approaches referenced above. In this respect, our first model uses the decision of Vissers et al. [57] on the patient's admission dates in order to perform a contribution margin-based big bucket scheduling decision, detailing on what day which clinical activities of admitted patients will take place. The second model generalizes the first model by additionally deciding on the day of admission. The decisions made by each of the models can then be employed as input for the detailed resource specific planning approaches in order to decide on more precise schedules and sequences for periods (such as hours or minutes) of less than a day.

Second, in terms of a deterministic or stochastic approach, our paper follows the aggregated approaches of Vissers et al. [57] and Conforti et al. [12]. The problems are modelled as a deterministic MIP using expected values for resource demands of clinical activities, for recovery times and resource capacities. The usage of expected values is well aligned with the big bucket scheduling problem, since deviations of the realized values from the expected values will be mitigated due to variance reduction effects, and there will be no disturbance of detailed time lines. However, with respect to the demand for bed capacity a deviation from the expected recovery times will disrupt the schedule. Hence, we embed our deterministic models in a rolling horizon approach in order to cope with this type of variance.

Third, with respect to the objective function our approach is the only one which considers contribution margin, a management accounting measure, and not an operational substitute measure, such as weighted utilization or weighted time.

Finally, with respect to the constraints, our models are, along with Vissers et al. [57], the only ones which consider all scarce clinical resources in an aggregated manner.

3 Model formulations

In the following, we present the two models for planning the patient flow, one with fixed and one with variable admission dates.

3.1 The patient flow problem with fixed admission dates

Patients and clinical pathways We consider a set of elective patients $\mathcal{P}$ that have to be treated in the hospital. For each patient $p \in \mathcal{P}$ we know its clinical pathway which is a standardized, typically evidence based health care process (see van De Klundert et al. [54]). In our models we depict the clinical pathway of a patient as an activity-on-node graph in which the set of nodes represents the clinical activities and weighted arcs represent minimum time lags between clinical activities. Let $\mathcal{A}$ denote the set of all clinical activities to be performed and let $\mathcal{E}$ denote the set of all minimum time lags between clinical activities. A minimum time lag $(i, j) \in \mathcal{E}$ of weight $d_{i,j}^{\min} \in \mathbb{Z}_{\geq 0}$ stipulates that activity j has to be performed at least $d_{i,j}^{\min}$

periods later than activity i; a period in our model accounts for one day. In Section 3.3 we provide a detailed example of two patients and their corresponding clinical pathways. Given the graph $(\mathcal{A}, \mathcal{E})$, the admission period α_p and the discharge activity ϕ_p for each patient $p \in \mathcal{P}$, we calculate for each activity $i \in \mathcal{A}$ the earliest period E_i and the latest period L_i in which the activity has to be performed with longest path methods (see, for example, Neumann et al. [35]; details of the calculation are given in Section 4.1). Let $\mathcal{W}_i := \{E_i, E_i + 1, \ldots, L_i\}$ denote the time window of activity i and let $\mathcal{T}$ be the set of periods (days) within the planning horizon. The length of the planning horizon $|\mathcal{T}|$ is set to the maximum latest period of all activities $\max_{i \in \mathcal{A}} L_i$ which equals the latest discharge period of all patients.

Hospital resources Scarce hospital resources are depicted by two sets, day resources and overnight resources. The set of day resources $\mathcal{R}^d$ depicts all resources which are available during regular working days, such as computer tomography scanner (CT), magnetic resonance imaging scanner (MRI) and surgery room. Each day resource $k \in \mathcal{R}^d$ has a capacity $R_{k,t}$ in period $t \in \mathcal{T}$. The capacity demand of activity $i \in \mathcal{A}$ on day resource $k \in \mathcal{R}^d$ is $r_{i,k}$. $R_{k,t}$ and $r_{i,k}$ are typically measured in minutes. The capacity demand $r_{i,k}$ of clinical activities ranges between several minutes and several hours, whereas the capacity of day resources is, in the case of a single shift, 8 hours. The set of overnight resources is depicted by $\mathcal{R}^n$. Overnight resources represent beds available at specialties, for example, the number of beds available at the surgical specialty. The number of beds available at overnight resource $k \in \mathcal{R}^n$ in period $t \in \mathcal{T}$ is $R_{k,t}$. If patient $p \in \mathcal{P}$ stays in the hospital from day t to day $t + 1$, he requires one bed from the overnight resource b_p in period t. With $\mathcal{R} = \mathcal{R}^d \cup \mathcal{R}^n$ we denote the set of all resources which includes day and overnight resources.

DRG and contribution margin We consider a DRG system in which, given a patient's DRG and his admission day, the hospital receives a specific revenue which depends on the DRG and the discharge date and thus the length of stay. Figures 1(a)–(b) show two length of stay and DRG dependent revenue functions for a specific DRG (based on the German DRG system). The revenue (dotted line) is constant if the length of stay (LOS) is between a low LOS trim point and a high LOS trim point. The trim points, 2 and 9 in Figure 1, represent the length of stay range for the majority of the patients with a "normal" healing process. For a length of stay within this range the hospital receives a constant revenue. In the case of a length of stay below the low LOS trim point there is a per-day reduction of the revenue. In this scenario, the health insurance company takes into account the fact that the hospital has lower costs than for the average patient with the same DRG. In contrast, if for medical reasons the length of stay of a patient is above the high LOS trim point, the hospital receives a per-day surcharge on top of the fixed revenue in order to compensate additional costs (see case (a) in Figure 1). However, if a length of stay above the high LOS trim point is not medically

justified, this surcharge will not be paid (see case (b) in Figure 1). In practice, a claims agent of the health insurance company scrutinizes the bill, examines the medical record and then makes a decision regarding the surcharge.
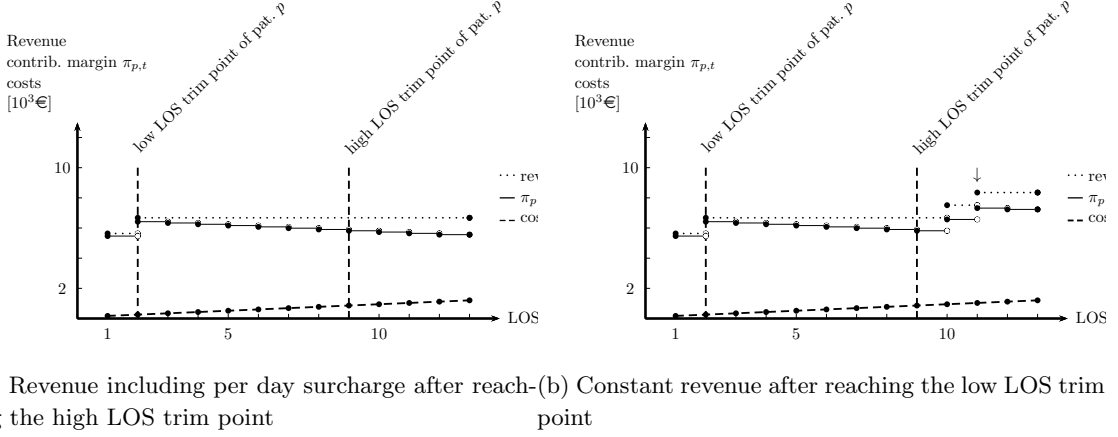

(a) Revenue including per day surcharge after reaching the high LOS trim point

(b) Constant revenue after reaching the low LOS trim point

Figure 1: Two example revenue, cost and contribution margin functions

Country	Low LOS trim point	High LOS trim point	Reference
Australia	yes	yes	Sharma [51]
Austria	yes	yes	Busse et al. [5]
Canada	no	yes	Botz [3]
Denmark	no	yes	Schreyögg et al. [47]
England	no	yes	Schreyögg et al. [47]
France	yes	yes	Schreyögg et al. [47]
Germany	yes	yes	Schreyögg et al. [47]
Italy	no	yes	Schreyögg et al. [47]
Spain	no	yes	Schreyögg et al. [47]
Switzerland	yes	yes (2)	Zaugg et al. [58]
USA	no	yes	Rogers et al. [43]

Table 2: DRG-based reimbursement policies of different countries

Table 2 provides an overview of the DRG-based reimbursement policies of different countries. The table reveals that high LOS trim points exist in all listed countries while low LOS trim points are only in place in some of the countries. A reason for countries not to employ low LOS trim points is to provide an incentive for short-stay visits or day-cases (for example in England, see Schreyögg et al. [47]). The parameter of the revenue function, low and high LOS trim points, fixed revenue as well as per day reduction and addition are DRG specific (see for example Porter and Guth [40]). Since we measure the length of stay in discrete days, the revenue function is discrete as well. From the revenue, we subtract the variable cost not associated with clinical procedures, which are essentially costs for meals, cleaning and so on

(dashed line in Figure 1), in order to obtain the contribution margin $\pi_{p,t-\alpha_p}$ for patient p with admission day α_p, discharge day t and thus length of stay $t - \alpha_p$ (solid line in Figure 1). Note that we do not subtract the cost for clinical procedures because, in accordance with the clinical pathway, the latter will be undertaken in any case and are therefore not relevant for the decision, whereas the patients' length of stay is a function of the scheduling decision. Also note that the time-constant DRG variant is a special case of the time-variable variant and, hence, our model can be applied directly to the latter.

Decision variables and model formulation   Using the binary variables

$$
x_{i,t} = \begin{cases} 1, & \text{if clinical activity } i \in \mathcal{A} \text{ is done at day } t \in \mathcal{W}_i \\ 0, & \text{otherwise} \end{cases}
$$

originally introduced by Pritsker et al. [41], we can model the patient flow problem with fixed admission date (PFP–FA) as follows:

$$
\text{Maximize} \quad z = \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{W}_{\phi_p}} \pi_{p,t-\alpha_p} \cdot x_{\phi_p,t} \tag{1}
$$

subject to

$$
\sum_{t \in \mathcal{W}_j} t \cdot x_{j,t} \geq \sum_{t \in \mathcal{W}_i} t \cdot x_{i,t} + d_{i,j}^{\min} \qquad \forall (i,j) \in \mathcal{E} \tag{2}
$$

$$
\sum_{i \in \mathcal{A}: t \in \mathcal{W}_i} r_{i,k} \cdot x_{i,t} \leq R_{k,t} \qquad \forall k \in \mathcal{R}^d, t \in \mathcal{T} \tag{3}
$$

$$
\sum_{p \in \mathcal{P}: b_p = k, t \geq \alpha_p} \left(1 - \sum_{\tau = E_{\phi_p}}^{\min\{t, L_{\phi_p}\}} x_{\phi_p, \tau} \right) \leq R_{k,t} \qquad \forall k \in \mathcal{R}^n, t \in \mathcal{T} \tag{4}
$$

$$
\sum_{t \in \mathcal{W}_i} x_{i,t} = 1 \qquad \forall i \in \mathcal{A} \tag{5}
$$

$$
x_{i,t} \in \{0,1\} \qquad \forall i \in \mathcal{A}, t \in \mathcal{W}_i \tag{6}
$$

The objective function (1) maximizes the contribution margin of all patients who have to be admitted to the hospital. Constraints (2) ensure minimum time lags between clinical activities. Note that for $d_{i,j}^{\min} = 0$ activities i and j can be performed on the same day. This might be for example the case for two diagnostic activities. Constraints (3) depict the limited capacity of day resources. For each resource and day, the capacity demand of all activities performed at that day must not exceed resource capacity. Note that the dimension of resource capacity can be, for example, time or slots in a master surgical schedule. Constraints (4) denote the resource constraints for overnight resources. For patient p, the bed allocation starts with the fixed admission date α_p and the discharge activity ϕ_p releases the bed. Constraints (5) ensure

that each activity is assigned to exactly one period of its time window. Variable definitions are provided in (6).

Table 3 provides an example of the bed allocation with a planning horizon of $\mathcal{T} := \{1, \ldots, 7\}$ days, overnight resource $\mathcal{R}^n := \{1\}$ and a single patient $\mathcal{P} := \{1\}$ with admission date $\alpha_1 = 1$ requiring overnight resource $b_1 = 1$. The first row provides the decision variables $x_{\phi_1, t}$ for discharge activity ϕ_1 of patient 1 which has to take place in exactly one period within the time window $\mathcal{W}_{\phi_1} := \{5, 6, 7\}$. As can be seen, the PFP–FA selects the discharge period 6 by setting $x_{\phi_1, 6} = 1$. The second row gives the left hand side of constraints (4) for patient 1. The patient requires a bed starting with the night between days 1 and 2 until the night between days 5 and 6.

t		1	2	3	4	5	6	7
$x_{\phi_1, t}$		–	–	–	–	0	1	0
$\displaystyle\sum_{b_1=1, t \geq 1} \left(1 - \sum_{\tau = E_{\phi_1}}^{\min\{t, L_{\phi_1}\}} x_{\phi_1, \tau} \right)$		1	1	1	1	1	0	0

Table 3: Bed allocation for the PFP–FA

In contrast to resource-constrained project scheduling (see in Pritsker et al. [41]), the PFP–FA does not undertake detailed scheduling in which an activity's starting and ending time is precisely planned. Instead, the PFP–FA does aggregated planning by assigning activities to days instead of deciding on activity start and finish times during the day. This aggregated view can be illustrated by the relation between the length of periods and the duration of activities. In resource-constrained project scheduling, we typically have activity durations which are multiples of a period. Conversely, in the case of our model, activity durations are between 10 minutes and several hours, whereas the length of one period is a working day with, in the case of one shift, 8 hours. The approach of our model can thus be compared to assembly line balancing (see for example Boysen et al. [4]) in which tasks are assigned to stations but no sequencing of tasks within stations is undertaken.

Note that model (1)–(6) does not take into account a master surgery schedule (MSS) in which OR capacity is assigned on the basis of a day or half-days to specialties but assumes that surgeries from any specialty can be undertaken any day. However, the model can be generalized straightforwardly in order to take MSSs into account. For this, constraints (3) have to be defined not for the aggregated single day resource "operating theatre" but for each specialty. In consequence, the number of constraints (3) increases.

The number of binary decision variables is $\sum_{i \in \mathcal{A}} |\mathcal{W}_i|$ while the number of constraints is $|\mathcal{E}| + |\mathcal{R}| \cdot |\mathcal{T}| + |\mathcal{A}|$. For a realistic problem with a planning horizon of $|\mathcal{T}| = 28$ days, $|\mathcal{P}| = 150$ patients, $|\mathcal{R}| = 10$ resources, 3 activities per patient, time window size of $|\mathcal{W}_i| = 3$ days for each activity $i \in \mathcal{A}$ and $|\mathcal{E}| = 300$ precedence relations, we have 1,350 binary decision variables

9

and 1,030 constraints. Tables 9–11 in Section 4 provide the problem sizes of real-world test instances.

3.2 The patient flow problem with variable admission dates

In the case of elective patients, the admission date is typically negotiated between the patient or the referring physician and the hospital. Hence, α_p is no longer a parameter. Instead, we introduce for each patient p admission activity σ_p which has to be assigned by variables $x_{\sigma_p,t}$ to exactly one admission day t within time window $\mathcal{W}_{\sigma_p}$. Note that in the case of variable admission dates, the contribution margin function $\pi_{p,t}$ for patient $p \in \mathcal{P}$ is defined for lengths of stay $t \in \mathcal{L}_p$, calculated by subtracting the admission date from the discharge date, formally given by

$$\sum_{t \in \mathcal{W}_{\phi_p}} t \cdot x_{\phi_p,t} - \sum_{t \in \mathcal{W}_{\sigma_p}} t \cdot x_{\sigma_p,t}. \tag{7}$$

In order to ease the readability of the model, we introduce the binary variables

$$y_{p,t} = \begin{cases} 1, & \text{if patient } p \in \mathcal{P} \text{ has a LOS of } t \in \mathcal{L}_p \text{ days} \\ 0, & \text{otherwise.} \end{cases}$$

We can now formulate the patient flow problem with variable admission dates (PFP–VA) as follows:

$$\text{Maximize} \quad z = \sum_{p \in \mathcal{P}} \sum_{t \in \mathcal{L}_p} \pi_{p,t} \cdot y_{p,t} \tag{8}$$

subject to (2)–(3), (5)–(6) and

$$\sum_{p \in \mathcal{P}:b_p=k} \left(\sum_{\tau=E_{\sigma_p}}^{\min\{t,L_{\sigma_p}\}} x_{\sigma_p,\tau} - \sum_{\tau=E_{\phi_p}}^{\min\{t,L_{\phi_p}\}} x_{\phi_p,\tau} \right) \le R_{k,t} \qquad \forall k \in \mathcal{R}^n, t \in \mathcal{T} \tag{9}$$

$$\sum_{t \in \mathcal{W}_{\phi_p}} t \cdot x_{\phi_p,t} - \sum_{t \in \mathcal{W}_{\sigma_p}} t \cdot x_{\sigma_p,t} = \sum_{t \in \mathcal{L}_p} t \cdot y_{p,t} \qquad \forall p \in \mathcal{P} \tag{10}$$

$$\sum_{t \in \mathcal{L}_p} y_{p,t} = 1 \qquad \forall p \in \mathcal{P} \tag{11}$$

$$y_{p,t} \in \{0,1\} \qquad \forall p \in \mathcal{P}, t \in \mathcal{L}_p \tag{12}$$

Objective function (8) maximizes the contribution margin of all patients. Overnight resource constraints (9) have been modified in order to take into account flexible admission dates. Constraints (10) calculate the LOS as the difference between the start of the discharge activity and the start of the admission activity. Constraints (11) ensure that for each patient exactly one LOS is selected and the additional LOS decision variables are defined by (12).

Table 4 provides an extension of the example in Table 3 where, additionally, we assume for patient 1 admission time window $\mathcal{W}_{\sigma_1} := \{1,2,3\}$. The first row provides the decision

variables for the admission activity σ_1 of patient 1 as well as the decision of PFP–VA to admit patient 1 in the first period ($x_{\sigma_1,1} = 1$). The second row gives the time window for the discharge activity as well as the decision of PFP–VA to discharge patient 1 in period 5 ($x_{\phi_1,5} = 1$). The third row provides the left hand side of the bed allocation constraints (9).

$t \in \mathcal{T}$	1	2	3	4	5	6	7
$x_{\sigma_1,t}$	1	0	0	–	–	–	–
$x_{\phi_1,t}$	–	–	–	–	1	0	0
$\displaystyle\sum_{b_1=1}\left(\sum_{\tau=E_{\sigma_1}}^{\min\{t,L_{\sigma_1}\}} x_{\sigma_1,\tau} - \sum_{\tau=E_{\phi_1}}^{\min\{t,L_{\phi_1}\}} x_{\phi_1,\tau}\right)$	1	1	1	1	0	0	0

Table 4: Bed allocation for the PFP–VA

The number of binary decision variables of model PFP–VA is $\sum_{i\in\mathcal{A}}|\mathcal{W}_i| + \sum_{p\in\mathcal{P}}|\mathcal{L}_p|$ and the number of constraints is $|\mathcal{E}| + |\mathcal{R}| \cdot |\mathcal{T}| + 2 \cdot |\mathcal{P}| + |\mathcal{A}|$. Thus, PFP–VA is considerably larger than PFP–FA. For example, consider the problem instance in Section 3.1 with an additional admission activity per patient and $|\mathcal{L}_p| = 3$ possible LOSs for each patient $p \in \mathcal{P}$. Then, the number of precedence relations is $|\mathcal{E}| = 450$, the number of binary decision variables is 1,800 and the number of constraints is 1,630. For real-world instances, again, see Tables 9–11 in Section 4.

3.3 An example of the patient flow problem with fixed and variable admission dates

In what follows we give an example with two patients. Patient (pat.) 1 has the DRG I53Z (spinal disc surgery) and patient 2 has the DRG B04D (extracranial surgery). Consider $\mathcal{T} := \{1, \ldots, 7\}$ time periods, activities $i \in \mathcal{A}$ and resources $\mathcal{R}^d := \{1, 2\}$ and $\mathcal{R}^n := \{3\}$ as shown in Table 5. The overnight resource requirements are $b_1 = b_2 = 3$. Clinical pathways and resource requirements for day resources are shown in Figure 2.

$i \in \mathcal{A}$		
PFP–FA	PFP–VA	Description
–	1	Admission of pat. 1
2	2	Spine CT for pat. 1
3	3	Spinal surgery of pat. 1
4	4	Discharge of pat. 1
–	5	Admission of pat. 2
6	6	Arteriography for pat. 2
7	7	Stent implantation of pat. 2
8	8	Discharge of pat. 2

(a)

$k \in \mathcal{R}$	Description	$R_{k,t}$
1	Radiology unit	30 min. on workdays, 0 otherwise
2	Operating theater	100 min. on workdays, 0 otherwise
3	Surgical ward	2 beds on workdays, 1 otherwise

(b)

Table 5: Activities (a) and resources (b)

Figure 2: Clinical pathways for the PFP–FA (a) and the PFP–VA (b)

In Figure 2(a), $r_{7,2} = 100$ denotes that activity 7 (stent implantation for patient 2) requires 100 minutes from day resource 2 which is the operating theater. The minimum time lag $d_{3,4}^{\min} = 4$ between activity 3 and 4 in Figure 2(b) denotes that at least four days of recovery time have to pass between the surgery and the discharge of patient 1. For the PFP–FA let the admission days for both patients be $\alpha_1 = \alpha_2 = 1$. Table 6 provides earliest and latest time periods to schedule the activities as obtained by longest path calculation.

$i \in \mathcal{A}$	2	3	4	6	7	8	$i \in \mathcal{A}$	1	2	3	4	5	6	7	8
E_i	1	1	5	1	1	5	E_i	1	1	1	5	1	1	1	5
L_i	3	3	7	3	3	7	L_i	3	3	3	7	3	3	3	7

(a)   (b)

Table 6: Earliest and latest periods for the PFP–FA (a) and the PFP–VA (b)

Table 7 shows the contribution margins for the two patients. Note that with increasing t, this margin decreases.

t	4	5	6
$\pi_{1,t}$	3,772.67	3,711.80	3,650.94
$\pi_{2,t}$	3,498.41	3,436.15	3,373.90

Table 7: Contribution margin (in €) for the PFP–FA and the PFP–VA

Solving the PFP–FA and the PFP–VA with the given parameters, we obtain the optimal solutions presented in Figures 3(a)–(b).
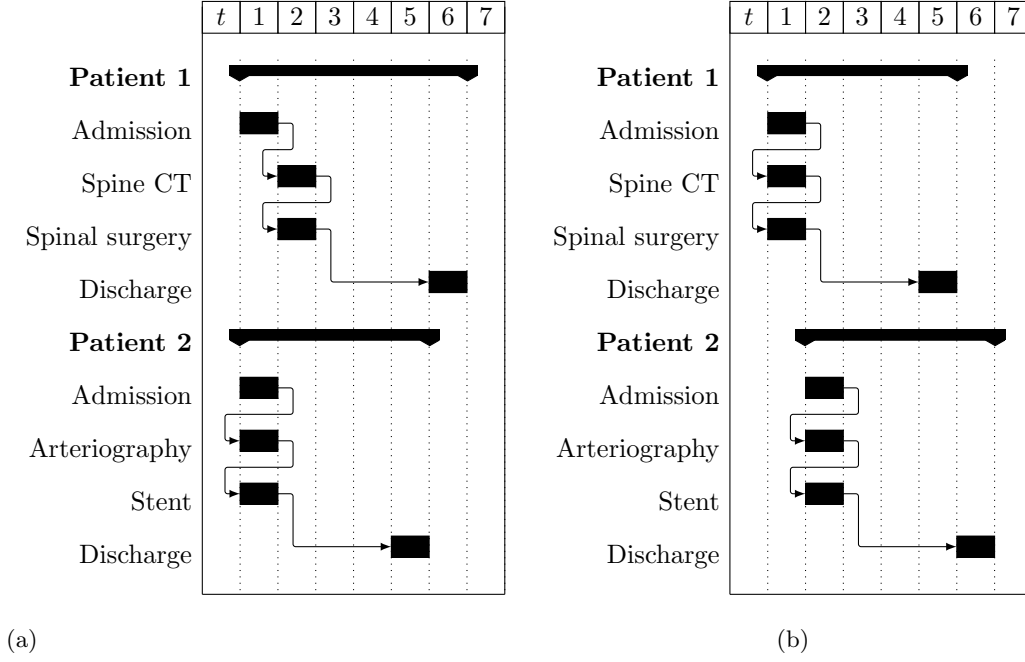
Figure 3: Optimal solution for the PFP–FA (a) and the PFP–VA (b)

The optimal objective function value of the PFP–FA and the PFP–VA is € 7,210.21 and € 7,271.08, respectively. Note that in the case of fixed admission dates, the length of stay of patient 1 is one day longer than in the case of variable admission dates. The admission date and the discharge date of patient 2 are postponed by one day in the case of flexible admission dates, resulting in the same length of stay as in the fixed case.

4 Computational study

In the following, a computational and economic analysis of scheduling the hospital-wide patient flow is provided.

4.1 Data and instance generation

We evaluated our models on data from the county hospital Erding, a 400-bed sized hospital in the vicinity of Munich, Germany. We joined the hospital data with data from the German institute for the reimbursement in hospitals (see Schreyögg et al. [48]). The latter contains the information on the DRG attributes such as low and high LOS trim point, fixed revenue as well as per day reduction and addition. Following suggestions from the hospital's management, the 18 DRGs shown in Table 8 were selected because at the time a patient contacts the hospital for admission, these DRGs can be classified with sufficient accuracy. The table provides the number of patients for each DRG (n), the classification accuracy (Acc.), i.e. the probability of assigning the correct DRG at admission and the surgery lead time. The surgery lead time is the minimum number of days required for medical activities prior to surgery. Examples

13

are laboratory tests which have to be made prior to an incision or the time required for pre-medication to take effect in the patient's metabolism before surgery. We depict surgery lead time with a minimum time lag between the admission activity and the surgery activity. For the year 2008, the 18 selected DRGs represent 11.6% of all admitted patients. In the following, we assume that these DRGs can be classified with high accuracy at the time the patient or the referring physician contacts the hospital. For the PFP–FA, we set for each patient $p \in \mathcal{P}$ the admission date α_p according to the realized admission date as given by the hospital data. In the case of the PFP–VA the admission time window $\mathcal{W}_{\sigma_p}$ is set according to the realized admission week. For example, if a patient was admitted on Wednesday, then for the PFP–FA we set α_p to that specific Wednesday whereas for the PFP–VA we require that the patient has to be admitted between Monday and Friday of that specific week. Only 3.8% of the patients required capacity on the ICU and of them 42.8% left the ICU within 24 hours. For the sake of simplicity, we therefore model the ICU as a day resource. Pre- and post-clinical treatments are not considered. Resource capacity $R_{k,t}$ is determined by using the entire capacity of resource $k \in \mathcal{R}$ on day $t \in \mathcal{T}$ minus capacity allocated to patients who do not have the selected DRGs. The order of activities $(i,j) \in \mathcal{E}$ is set according to the order of time stamps in the hospital data. For the capacity demand $r_{i,k}$ of clinical activity i on day resources k we use averages (avg.) of the observed distributions. As already stated above, this approach is robust because, first, the capacity demand of activities is a fraction of the available capacity per day, and, second, we benefit from variance reduction effects when summing up the capacity demands of activities assigned to the resource on the same day. The recovery time of each patient is determined by the difference between the time stamps of the discharge activity and its immediately preceding activity in the clinical pathway. For example, for patient 2 in Figure 2(b) activity 7 (stent implementation) immediately precedes activity 8 (discharge). Note that a surgery activity is not always the immediate predecessor of the discharge activity. For example, for DRG I68C in Table 8 a pain therapy activity is the immediate predecessor of the discharge activity.

DRG	Description	n	Acc. [%]	Surgery lead time [days]
B04D	Extracranial surgery	47	96.6	1
C08B	Cataract surgery	61	100	0
D30B	Tonsillectomy	45	100	0
D38Z	Nose surgery	50	89.4	0
F39B	Vein stripping	138	89.8	0
G24Z	Hernia repair	120	98.4	0
H08B	Laparoscopic cholecystectomy	169	100	0
I44B	Knee arthroplasty	62	93.1	1
I53Z	Spinal disc surgery	139	98.3	0
I68C	Spinal disc therapy (age > 55 years)	260	87.1	0
I68D	Spinal disc therapy (age < 56 years)	165	81.6	0
K12Z	Thyroid surgery	28	88.2	0
L06B	Non-complex urinary bladder incision	70	90.4	0
L20C	Complex urinary bladder incision	132	95.2	0
L64A	Bladder stone (age > 75 years)	127	90.8	0
M02Z	Prostate resection	54	100	0
N21Z	Hysterectomy	52	89.7	1
N25Z	Abrasio	51	82.4	0

Table 8: Statistics for the selected DRGs

Time windows for activities are determined by using longest path calculation (see Neumann et al. [35]). More precisely, in the case of the PFP–FA we set the earliest period of the first activity of a patient p equal to the admission date α_p. We then calculate earliest periods for all other activities of the patient using longest path methods. In doing so, the minimum surgery lead time between an activity preceding surgery and the surgery activity itself has to be taken into account for some surgeries (see column "Surgery lead time" in Table 8). In order to determine the latest periods L_i, we set $L_{\phi_p} = E_{\phi_p} + w$ using the time window variation parameter w and calculate the latest periods for all other activities of patient p. For the PFP–VA we set the earliest period of the admission activity σ_p to the Monday of the week in which the patient was admitted. We then calculate earliest periods as for the PFP–FA. For the latest period of the discharge activity we set $L_{\phi_p} = E_{\phi_p} + 4 + w$. The "4" is added in order to allow for a time window variation parameter of $w = 0$ which ensures that the patient is admitted between Monday and Friday. The time window variation parameter w has been varied between 0 and 6 (see Figure 4).

In total, we generated 12 test instances. Instance 1, 2, ..., 12 represents all elective patients with the DRGs as provided in Table 8 admitted in month 1, 2, ..., 12 of 2008. Table 9 provides the number of patients for each month as well as other key figures when the time window parameter w is set to $w = 4$ for the PFP–FA and $w = 1$ for the PFP–VA.

Instance	$\lvert\mathcal{T}\rvert$		$\lvert\mathcal{P}\rvert$	$\lvert\mathcal{A}\rvert$		$\lvert\mathcal{E}\rvert$		$\frac{1}{\lvert\mathcal{A}\rvert}\sum\limits_{i\in\mathcal{A}}\lvert\mathcal{W}_i\rvert$	
	FA	VA		FA	VA	FA	VA	FA	VA
1	53	52	179	453	632	274	453	5.6	5.3
2	45	45	142	356	498	214	356	5.6	5.5
3	46	46	145	335	480	190	335	5.3	5.6
4	47	45	185	449	634	264	449	5.4	5.2
5	53	52	137	350	487	213	350	5.5	5.5
6	51	48	147	377	524	230	377	5.7	5.2
7	47	46	141	347	488	206	347	5.4	5.3
8	49	50	114	286	400	172	286	5.4	5.6
9	44	42	136	340	476	204	340	5.5	5.2
10	45	46	154	379	533	225	379	5.5	5.5
11	44	44	153	363	516	210	363	5.4	5.6
12	33	34	137	320	457	183	320	5.4	5.6
avg.	46.4	45.8	147.5	362.9	510.4	215.4	362.9	5.5	5.4

Table 9: Sizes of the problem instances

As can be seen in the last two columns, this choice of w results in time windows of comparable size of the two models which are abbreviated as "FA" and "VA" in the table. The available capacity of the hospital resources has been adjusted by those patients which are not considered in the instances. These are non-elective patients and elective patients with DRGs different from the ones given in Table 8. Table 10 provides the daily capacity of the 12 day and 3 overnight resources before subtracting the expected demand of unplanned patients. For example, the daily operating theater capacity of 3,600 minutes comes from 6 operating rooms, each one opens from 8 a.m. until 6 p.m. on Monday through Friday.

$k \in \mathcal{R}$	Description	$R_{k,t}$
1–9	Biopsy, endoscopy, functional test, punction, ultrasound, CT, arteriography, nuclear medicine, MRI resources	480 min. on workdays, 0 otherwise
10	Operating theater	3,600 min. on workdays, 0 otherwise
11	Physical therapy	1,900 min. on workdays, 0 otherwise
12	ICU	40 beds on workdays, 90% of the workday capacity otherwise
13	Ear, nose and throat	155 beds on workdays, 90% of the workday capacity otherwise
14	Internal medicine	115 beds every day, 90% of the workday capacity otherwise
15	Surgical specialty	95 beds every day, 90% of the workday capacity otherwise

Table 10: Resources and capacities

4.2 Computation time analysis for the static approaches

Table 11 shows the number of variables and constraints as well as the computation times. All computations were performed on a 2 GHz PC (Intel Core2 Duo T7250) with 4 GB RAM running Windows 7 operating system. The models were coded in Java in an ILOG Concert environment. The solver used was ILOG CPLEX 12.2 (64 bit).

Instance	#Var.		#Constr.		Computation time [s]	
	FA	VA	FA	VA	FA	VA
1	2,558	5,031	1,701	2,402	0.1	0.8
2	1,995	4,135	1,387	1,955	0.1	0.3
3	1,778	4,124	1,360	1,940	0.2	0.4
4	2,435	4,951	1,603	2,313	0.2	0.9
5	1,917	4,036	1,495	2,028	0.2	0.5
6	2,165	4,102	1,519	2,062	0.1	0.5
7	1,886	3,913	1,399	1,948	0.1	0.4
8	1,553	3,350	1,307	1,778	0.1	0.3
9	1,877	3,718	1,340	1,854	0.2	0.3
10	2,067	4,451	1,433	2,064	0.1	0.4
11	1,945	4,430	1,386	1,998	0.1	0.6
12	1,742	3,942	1,135	1,698	0.1	0.4
avg.	1,993.2	4,181.9	1,422.1	2,003.3	0.1	0.5

Table 11: Number of decision variables, constraints and solutions times

Each instance was optimally solved in less than a second by each of the two models, with PFP–VA requiring solution times of factor 2 or more than the PFP–FA.

4.3 Economic analysis for the static approaches

Table 12 provides the absolute and relative improvement of the contribution margins obtained by the models when compared to the hospital data. The hospital does not have a formal approach for dealing with the patient flow problem. Instead, it employs a heuristic procedure for deciding on the discharge date and thus patients' LOS. The procedure works as follows. First, the medical criteria always have priority over the economic criteria. That is, a patient is not discharged until fully recovered. As soon as a patient has sufficiently recovered, the discharge decision is based on the marginal contribution margin. The patient is kept in the hospital as long as the marginal contribution margin is increasing and is discharged as soon as the marginal contribution margin is decreasing.

Instance	Improvement [€]		Improvement [%]	
	PFP–FA	PFP–VA	PFP–FA	PFP–VA
1	7,252.3	16,976.5	2.3	5.3
2	8,563.9	15,507.5	3.0	5.5
3	13,352.1	16,971.6	5.2	6.6
4	10,470.2	16,497.7	3.0	4.7
5	6,296.4	16,748.8	2.5	6.5
6	8,932.3	15,146.8	3.0	5.1
7	7,577.6	17,932.8	3.2	7.5
8	7,000.0	13,779.2	3.1	6.0
9	12,275.7	15,599.7	4.7	6.0
10	10,969.7	15,749.3	3.9	5.6
11	10,530.4	14,194.3	3.8	5.1
12	9,605.1	14,054.0	4.1	6.0
avg.	9,402.1	15,763.2	3.5	5.8

Table 12: Economic improvement for the static approaches

The average relative economic improvement for solving the test instances to optimality compared to the hospital solution is 3.5% and 5.8% when employing the fixed admission date model PFP–FA and the flexible admission date model PFP–VA, respectively. This is equivalent to a monthly increase of the contribution margin of € 9,402.1 for the PFP–FA and of € 15,763.2 for the PFP–VA.

4.4 Rolling horizon approach

In practice, there is considerable variance in the data and therefore hospitals cannot undertake a flow management of patients for a whole month in advance. Instead, they have to embed the model into a rolling horizon approach in which the planning of the patient flow is done every day with the most recent information available. This is also done in practice. Here,

the hospital undertakes daily allocation decisions for single resources, such as the operating theater and the wards. For example, each afternoon a decision is made about which of the patients will be considered for surgery on the next day. Therefore, we have embedded each of the two patient flow models in a rolling horizon approach in which each problem instance is solved each day instead of once a month. In each run, the most recent data is used. This is deterministic data which has become available at the day of planning, such as planned admissions in the PFP–FA, or stochastic data for which expected values are calculated. More precisely, we embedded the PFP–FA in the rolling horizon approach as follows. We solve the model each day with all remaining, i.e. not yet undertaken, clinical activities of patients admitted at that day or patients not yet discharged. The available resource capacity per day is calculated by subtracting the average week-day dependent demand of patients not considered in our planning approach (non elective patients and elective patients with DRGs different from the ones given in Table 8) from the overall capacity. For the first day of each planning horizon we subtract the observed capacity demand of all patients not considered in our model from the overall capacity. The duration of the recovery activities is calculated with Bayesian updating. We employed the approach of Tu et al. [53] which successfully used Bayesian methods to predict recovery times. We added the attribute DRG and so far realized recovery time to the attributes age, gender and type of surgery. When embedding the PFP–VA in the rolling horizon approach, we considered all patients which were admitted in the week under consideration instead of all patients which were admitted at the day of planning as in the PFP–FA. Thereby, we assumed that all requests of elective patients are known at the beginning of the week. Using data of the first half of 2008 to initialize the probability tables of the Bayesian classifier and to forecast available capacity, we run the rolling horizon approach for the third quarter of 2008. When calculating the activity time windows for the rolling horizon approach, we proceeded similarly to the static approach with the following modifications. First, for both models, the PFP–FA and the PFP–VA, we employ recovery times as calculated with the Bayesian classifier instead of using realized recovery times. Second, for calculating latest time periods we set the latest period of the discharge activity L_{ϕ_p} of patient p equal to the maximum DRG-dependent recovery time. With this approach we obtain time windows which are on average larger than in the static case. In order to make the results comparable we adjusted the time window variation parameter to $w = 6$ for the fixed admission and to $w = 5$ for the variable admission model leading to a negligible difference between the static and the rolling horizon approach of 0.1 and 0.5 days for the PFP–FA and the PFP–VA, respectively (see Table 9).

For the PFP–FA, when running the model for day t, we admit all patients with $\alpha_p = t$. For the PFP–VA, when running the model for day t, we allow the model to admit all patients which were admitted at that day or some day later in the week. For example, if t is a Monday,

the model schedules the admission activity σ_p of all patients in the hospital data admitted in that week on exactly one day between Monday and Friday. However, only the decision of patients scheduled on Monday will be fixed. If t is a Tuesday, the model schedules the admission activities of all patients in the hospital data admitted in that week who were not scheduled for Monday and so forth.

Table 13 gives the gain in contribution margin for the PFP–FA and the PFP–VA when implemented in a static vs a rolling horizon approach for the third quarter of 2008.

	Improvement [€]		Improvement [%]	
	PFP–FA	PFP–VA	PFP–FA	PFP–VA
Static	46,844.2	47,311.7	6.4	6.5
Rolling horizon	27,441.0	33,427.0	3.9	5.1

Table 13: Economic improvement of the static vs the rolling horizon approaches

Note that the improvement for the static approach is larger than the sum of the improvements for the instances $7 - 9$ (July – September) given in Table 12. This is due to the larger time windows which were used in the rolling horizon approach. From Table 13 one can see that the improvement of both models compared to the hospital solution is reduced for the more realistic rolling horizon approach. This holds true to a larger extent for PFP–FA. Nevertheless, both approaches and in particular PFP–VA still provide significant improvements compared to the hospital solution. The inferior results of the rolling horizon approach are due to incomplete information. More precisely, for the rolling horizon approach a forecast of the recovery times is used rather than deterministic recovery times. Furthermore, the stochastic resource availability for the first period of each planning run is also considered.

Influence of the time window size Figures 4(a)–(b) show the impact of the time window variation parameter w on the computation time and the relative improvement of PFP–FA and PFP–VA.
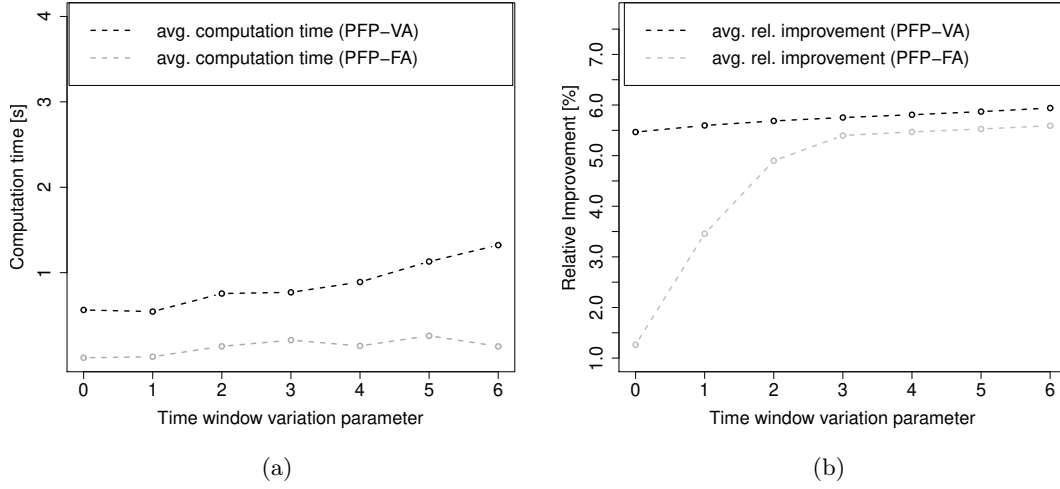
Figure 4: Computation time (a) and relative improvement (b) as a function of w

For increasing time window variation parameter values w the computation time and the economic improvement, compared to the hospital solution, are increasing as well. The increase of the computation time can be observed in particular for the PFP–VA, which considers, when compared to the PFP–FA, the same number of patients but larger networks due to more activities and more precedence constraints. The increase in the improvement is particularly strong for the PFP–FA. The impact on the improvement of the PFP–VA is only small because the latter uses, by definition, larger time window sizes due to the admission time window whereas the PFP–FA uses smaller ones. For both approaches, the PFP–FA and the PFP–VA, the marginal improvement beyond a time window variation parameter of $w = 3$ is small.

Length of stay and time admitted patients wait for surgery Figures 5(a)–(b) provide the frequency distribution for the time span between admission and surgery as well as the length of stay in the hospital.
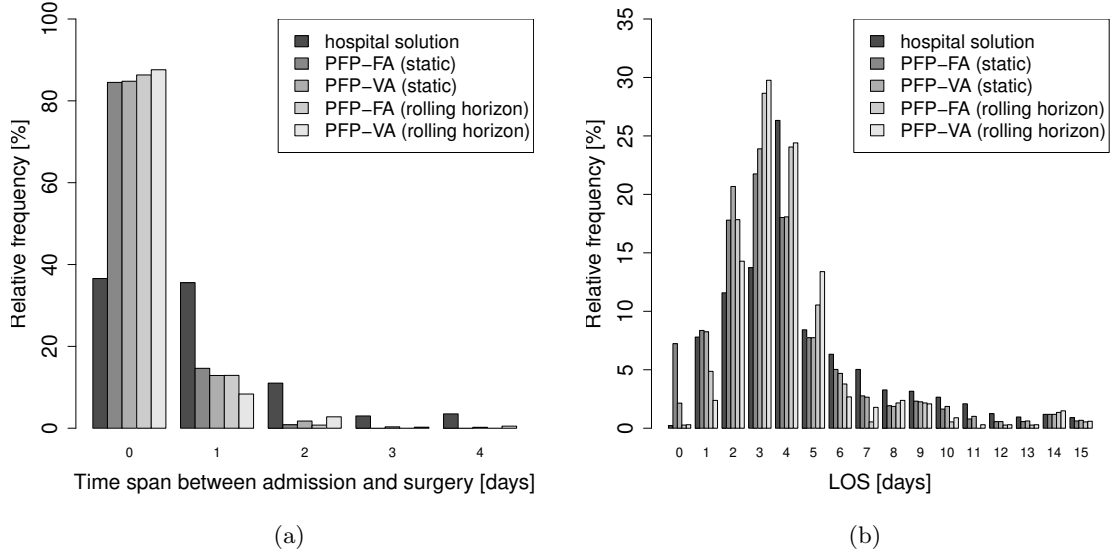
Figure 5: Time span between admission and surgery (a) and LOS (b)

Our models reduce the time span between admission and surgery compared to the hospital solution. In the hospital solution the difference between the surgery and the admission date is on average 1.8 days. In both, the solutions of the static PFP–FA and the PFP–VA, the average time span between admission and surgery is 0.2 days while it is 0.1 and 0.2 days for the rolling horizon PFP–FA and PFP–VA, respectively. Employing our models also leads to a reduction in the length of stay. Compared to the hospital solution with an average LOS of 6.5 days, the static PFP–FA and PFP–VA reduce the length of stay to 4.2 days. The rolling horizon implementation results in a slightly longer length of stay of 4.3 days for the PFP–FA and 4.6 days for the PFP–VA, respectively.

5   Conclusion

In this paper we have presented two discrete optimization models for the problem of scheduling elective patients hospital-wide with the objective to maximize the contribution margin. For the first model, we assumed a fixed admission date whereas in the second model this assumption is relaxed. We embedded both models into a rolling horizon approach in order to cope with stochastic data. Using real-world data from a mid-size hospital we performed an elaborate computational study in order to study the applicability and the impact of the models. The results show that the test instances can be solved to optimality within a very moderate time of a fraction of a second on average. The obtained results are clearly superior to the current solution of the hospitals which are derived with a simple heuristic. Analyzing our solutions we observed that the time span between admission and surgery and the length of stay in the hospital can be decreased substantially.

References

[1] V. Augusto, X. Xie and V. Perdomo. Operating theatre scheduling with patient recovery in both operating rooms and recovery beds. *Computers & Industrial Engineering*, 58(2): 231–238, 2010.

[2] S. Batun, B. T. Denton, T. R. Huschka and A. J. Schaefer. Operating room pooling and parallel surgery processing under uncertainty. *INFORMS Journal on Computing*, 23(2): 220–237, 2011.

[3] C. K. Botz. Principles for funding on a case mix basis: Construction of case weights (RIWs). *Healthcare Management Forum*, 4(4):22–32, 1991.

[4] N. Boysen, M. Fliedner and A. Scholl. A classification of assembly line balancing problems. *European Journal of Operational Research*, 183(2):674–693, 2007.

[5] R. Busse, A. Geissler, W. Quentin and M. Wiley. *Diagnosis-related groups in Europe.* McGraw-Hill, Berkshire, 2011.

[6] B. Cardoen and E. Demeulemeester. A decision support system for surgery sequencing at UZ Leuven's day-care department. *International Journal of Information Technology & Decision Making*, 10(3):435–450, 2011.

[7] B. Cardoen, E. Demeulemeester and J. Beliën. Optimizing a multiple objective surgical case sequencing problem. *International Journal of Production Economics*, 119(2):354–366, 2009.

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

[9] S. Ceschia and A. Schaerf. Local search and lower bounds for the patient admission scheduling problem. *Computers & Operations Research*, 38(10):1452–1463, 2011.

[10] S. Chaabane, N. Meskens, A. Guinet and M. Laurent. Comparison of two methods of operating theatre planning: Application in Belgian hospital. *Journal of Systems Science and Systems Engineering*, 17(2):171–186, 2008.

[11] V. S. Chow, M. L. Puterman, N. Salehirad, W. Huang and D. Atkins. Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. *Production and Operations Management*, 20(3):418–430, 2011.

[12] D. Conforti, F. Guerriero, R. Guido, M. M. Cerinic and M. Conforti. An optimal decision making model for supporting week hospital management. *Health Care Management Science*, 14(1):74–88, 2011.

[13] P. Demeester, W. Souffriau, P. De Causmaecker and G. Vanden Berghe. A hybrid tabu search algorithm for automatically assigning patients to beds. *Artificial Intelligence in Medicine*, 48(1):61–70, 2009.

[14] B. Denton and D. Gupta. A sequential bounding approach for optimal appointment scheduling. *IIE Transactions*, 35(11):1003–1016, 2003.

[15] B. Denton, J. Viapiano and A. Vogl. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science*, 10(1):13–24, 2007.

[16] B. Denton, A. Miller, H. Balasubramanian and T. Huschka. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research*, 58(4):802–816, 2010.

[17] S. Duckett. Casemix funding for acute hospital inpatient services in Australia. *The Medical Journal of Australia*, 169(8):17–21, 1998.

[18] H. Fei, C. Chu, N. Meskens and A. Artiba. Solving surgical cases assignment problem by a branch-and-price approach. *International Journal of Production Economics*, 112(1): 96–108, 2008.

[19] 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 Operations Research*, 166(1):91–108, 2009.

[20] H. Fei, N. Meskens and C. Chu. A planning and scheduling problem for an operating theatre using an open scheduling strategy. *Computers & Industrial Engineering*, 58(2): 221–230, 2010.

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

[22] A. Guinet and S. Chaabane. Operating theatre planning. *International Journal of Production Economics*, 85(1):69–81, 2003.

[23] E. W. Hans, M. van Houdenhoven and P. J. H. Hulshof. A framework for health care planning and control. In R. Hall, editor, *Handbook of Health Care Systems Scheduling*, volume 168 of *Springer International Series in Operations Research & Management Science*, pages 303–320. Springer, New York, 2011.

[24] P. J. H. Hulshof, N. Kortbeek, R. J. Boucherie, E. W. Hans and P. J. M. Bakker. Taxonomic classification of planning decisions in health care: A structured review of the state of the art in OR/MS. *Health Systems*, 1(2):129–175, 2012.

[25] A. Jebali, A. B. H. Alouane and P. Ladet. Operating rooms scheduling. *International Journal of Production Economics*, 99(1–2):52–62, 2006.

[26] S. V. Jerić and J. R. Figueira. Multi-objective scheduling and a resource allocation problem in hospitals. *Journal of Scheduling*, 15(5):513–535, 2012.

[27] S. E. Kimes. Yield management: A tool for capacity-considered service firms. *Journal of Operations Management*, 8(4):348–363, 1989.

[28] M. Lamiri, X. Xie, A. Dolgui and F. Grimaud. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research*, 185(3):1026–1037, 2008.

[29] M. Lamiri, X. Xie and S. Zhang. Column generation approach to operating theater planning with elective and emergency patients. *IIE Transactions*, 40(9):838–852, 2008.

[30] M. Lamiri, F. Grimaud and X. Xie. Optimization methods for a stochastic surgery planning problem. *International Journal of Production Economics*, 120(2):400–410, 2009.

[31] G. J. Lim, A. Mobasher, L. Kardar and M. J. Cote. Nurse scheduling. In R. Hall, editor, *Handbook of Health Care Systems Scheduling*, volume 168 of *Springer International Series in Operations Research & Management Science*, chapter 3, pages 31–64. Springer, New York, 2011.

[32] E. Marcon, S. Kharraja and G. Simonnet. The operating theatre planning by the follow-up of the risk of no realization. *International Journal of Production Economics*, 85(1): 83–90, 2003.

[33] I. Marques, M. Captivo and M. Vaz Pato. An integer programming approach to elective surgery scheduling. *OR Spectrum*, 34(2):407–427, 2012.

[34] D. Min and Y. Yih. An elective surgery scheduling problem considering patient priority. *Computers & Operations Research*, 37(6):1091–1099, 2010.

[35] K. Neumann, C. Schwindt and J. Zimmermann. *Project scheduling with time windows and scarce resources*. Springer, Berlin, 2nd Ed., 2003.

[36] I. Ozkarahan. Allocation of surgeries to operating rooms by goal programing. *Journal of Medical Systems*, 24(6):339–378, 2000.

[37] J. Patrick, M. Puterman and M. Queyranne. Dynamic multipriority patient scheduling for a diagnostic resource. *Operations Research*, 56(6):1507–1525, 2008.

[38] M. Persson and J. Persson. Analysing management policies for operating room planning using simulation. *Health Care Management Science*, 13(2):182–191, 2010.

[39] D. Pham and A. Klinkert. Surgical case scheduling as a generalized job shop scheduling problem. *European Journal of Operational Research*, 185(3):1011–1025, 2008.

[40] M. Porter and C. Guth. *Redefining German health care – Moving to a value-based system.* Springer, Heidelberg, 2012.

[41] A. A. B. Pritsker, L. Watters and P. M. Wolfe. Multiproject scheduling with limited resources: A zero-one programming approach. *Management Science*, 16(1):93–108, 1969.

[42] A. Riise and E. Burke. Local search for the surgery admission planning problem. *Journal of Heuristics*, 17(4):389–414, 2011.

[43] F. Rogers, T. Osler, S. Shackford, M. Cohen and L. Camp. Financial outcome of treating trauma in a rural environment. *The Journal of Trauma*, 43(1):65–73, 1997.

[44] B. Roland, C. Di Martinelly, F. Riane and Y. Pochet. Scheduling an operating theatre under human resource constraints. *Computers & Industrial Engineering*, 58(2):212–220, 2010.

[45] K. Schimmelpfeng, S. Helber and S. Kasper. Decision support for rehabilitation hospital scheduling. *OR Spectrum*, 34(2):461–489, 2012.

[46] R. Schmidt, S. Geisler and C. Spreckelsen. Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources. *BMC Medical Informatics and Decision Making*, 13(3):1–37, 2013.

[47] J. Schreyögg, T. Stargardt, O. Tiemann and R. Busse. Methods to determine reimbursement rates for diagnosis related groups (DRG): A comparison of nine European countries. *Health Care Management Science*, 9(3):215–223, 2006.

[48] J. Schreyögg, O. Tiemann and R. Busse. Cost accounting to determine prices: How well do prices reflect costs in the German DRG-system? *Health Care Management Science*, 9 (3):269–279, 2006.

[49] H.-J. Schütz and R. Kolisch. Approximate dynamic programming for capacity allocation in the service industry. *European Journal of Operational Research*, 218(1):239–250, 2012.

[50] J. A. Sepulveda. Prospective payment: A simulation model of management strategies. In *Proceedings of the 17th Conference on Winter Simulation*, pages 532–540. ACM, 1985.

[51] A. Sharma. Inter-DRG resource dynamics in a prospective payment system: A stochastic kernel approach. *Health Care Management Science*, 12(1):38–55, 2009.

[52] A. Testi and E. Tànfani. Tactical and operational decisions for operating room planning: Efficiency and welfare implications. *Health Care Management Science*, 12(4):363–373, 2009.

[53] J. V. Tu, S. B. Jaglal and C. D. Naylor. Multicenter validation of a risk index for mortality, intensive care unit stay, and overall hospital length of stay after cardiac surgery. *Circulation*, 91(3):677–684, 1995.

[54] J. van de Klundert, P. Gorissen and S. Zeemering. Measuring clinical pathway adherence. *Journal of Biomedical Informatics*, 43(6):861–872, 2010.

[55] P. T. Vanberkel, R. J. Boucherie, E. W. Hans, J. L. Hurink and N. Litvak. A survey of health care models that encompass multiple departments. *International Journal of Health Management and Information*, 1(1):37–69, 2010.

[56] B. Vijayakumar, P. Parikh, R. Scott, A. Barnes and J. Gallimore. A dual bin-packing approach to scheduling surgical cases at a publicly-funded hospital. *European Journal of Operational Research*, 224(3):583–591, 2013.

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

[58] M. Zaugg, A. Gattiker, C. Moneta and T. Reay. Effective leadership in hospital management: How a CEO can lead a hospital towards more "competitiveness". The GZO case in Switzerland. *Online Journal of International Case Analysis*, 3(1):1–24, 2012.