RESEARCH PAPER

Flexible Hospital-wide Elective Patient Scheduling

Daniel Gartner[a] and Rema Padman[b]

[a]School of Mathematics, Cardiff University, Cardiff, United Kingdom
[b]The H. John Heinz III College, Carnegie Mellon University, Pittsburgh, USA

ARTICLE HISTORY
Submitted: December 30, 2016; 1^{st} revision: August 2, 2017; 2^{nd} revision: October 31, 2018; 3^{rd} revision: February 11, 2019; accepted for publication: March 1, 2019

ABSTRACT

In this paper, we build on and extend Gartner and Kolisch (2014)'s hospital-wide patient scheduling problem. Their contribution margin maximizing model decides on the patients' discharge date and therefore the length of stay. Decisions such as the allocation of scarce hospital resources along the clinical pathways are taken. Our extensions which are modeled as a mathematical program include admission decisions and flexible patient-to-specialty assignments to account for multi-morbid patients. Another flexibility extension is that one out of multiple surgical teams can be assigned to each patient. Furthermore, we consider overtime availability of human resources such as residents and nurses. Finally, we include these extensions in the rolling-horizon approach and account for lognormal distributed recovery times and remaining resource capacity for elective patients. Our computational study on real-world instances reveals that, if overtime flexibility is allowed, up to 5% increase in contribution margin can be achieved by reducing length of stay by up to 30%. At the same time, allowing for overtime can reduce waiting times by up to 33%. Our model can be applied in and generalized towards other patient scheduling problems, for example in cancer care where patients may follow defined cancer pathways.

KEYWORDS
Admission Scheduling; Mathematical Programming; OR in health services; Patient Scheduling; Resource Allocation; Rolling-horizon approach

1. Introduction

The introduction of diagnosis-related groups (DRGs) in prospective payment systems has put pressure on hospitals to use resources efficiently (Schreyögg *et al.* (2006a); Sharma and Yu (2009)). Efficient hospital-wide patient scheduling challenges hospitals on multiple fronts. In order to be profitable, hospitals have to decide which elective patients to admit and when to schedule them on scarce resources given high fixed and low variable costs. Hospitals usually focus on two patient groups which are emergency and elective patients. Emergency patients visit a hospital unscheduled, in need of immediate care. Arrival times and resource allocation decisions for this type of patients are therefore out of the control of the hospital's operational offline decision making. Human resource planning is also restricted on staffing and shift scheduling decisions.

CONTACT D. Gartner and R. Padman. Email: gartnerd@cardiff.ac.uk and rpadman@cmu.edu

On the other hand, each individual who belongs to the group of elective patients is scheduled in advance.

Hospital-wide elective patient scheduling has been the focus of previous work and we build on the approach of Gartner and Kolisch (2014). In a nutshell, they developed two mathematical models. First, the patient flow problem with fixed admission dates and second, the patient flow problem with flexible admission dates. Both models were embedded in a rolling-horizon approach in order to measure how uncertainty affects the scheduling decisions on metrics such as length of stay and contribution margin.

Our approach extends their work in the following five ways. First, we include a decision if a patient is admitted or not. Second, our model decides on the assignment of a patient to one out of several specialties to account for multi-morbid patients. Third, one out of multiple surgical teams can be assigned to each patient. Fourth, we take into account overtime of human resources such as residents and nurses. Finally, we extend a rolling-horizon approach to account for lognormal distributed recovery times and remaining resource capacity for elective patients. We propose a mathematical programming formulation which is embedded into an extension of Gartner and Kolisch (2014)'s rolling horizon planning.

We assume that we can classify a large enough percentage of elective patients according to DRG which draws on the early DRG classification problem published by Gartner *et al.* (2015). The clinical pathway (CP) 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. CPs can be learned from transactional hospital data using Machine Learning approaches and has been the focus of Arnolds and Gartner (2018)'s strategic planning problem. Once a CP is assigned to a 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 DRG-based payment scheme which is in use for example in Germany (Schreyögg *et al.* (2006b)) as well as in other developed-world countries such as the U.S. The payment scheme is not necessarily linked to DRGs; all healthcare systems that have a length of stay and patient-dependent contribution margin function can use our model.

Since some data such as recovery times of patients with a length of multiple days and remaining resource capacity for elective patients can be stochastic, we embed our model in a rolling horizon approach. Our computational study on real-world instances reveals that, if overtime flexibility is allowed, up to 5% increase in contribution margin can be achieved by reducing length of stay by up to 30%. At the same time, allowing for overtime can reduce waiting times by up to 33%.

The remainder of this paper is structured as follows. In Section 2, we provide a review of mathematical models applied to patient scheduling and show to what extent our approach differs from previous and related work. Section 3 presents the mathematical model and illustrates it by means of an example followed by a description of how the model is embedded into an extension of Gartner and Kolisch (2014)'s rolling horizon approach in Section 4. In Section 5, we carry out an experimental study based on real-world data from our collaborating hospital and analyze the impact of our approaches on metrics such as contribution margin, waiting time and length of stay. Section 6 provides a discussion of assumptions made for the modelling and the rolling horizon planning. Section 7 closes the paper with conclusions.

2. Related Work

Patient scheduling is the process of assigning individual patients and/or patients' activities to time and/or healthcare resources (Gartner and Padman (2017b)) on the operational decision level. In contrast, appointment scheduling defines a blueprint of patients' appointments on a tactical level. While reviews that exclusively focus on patient scheduling are Gartner and Padman (2017b) and Gartner (2015), appointment scheduling problems have been reviewed by Hulshof *et al.* (2012), Ahmadi-Javid *et al.* (2017), Kazemian *et al.* (2017), Leeftink *et al.* (2018) as well as Marynissen and Demeulemeester (2019). In this section, we position our paper in the relevant patient scheduling literature.

One of the most relevant papers in the review of Gartner and Padman (2017b) is the paper from Gartner and Kolisch (2014) because we extend their work. The authors formulated a patient scheduling problem using binary programming with the objective to maximize contribution margin which is a function of length of stay dependent revenue and costs. The authors assume that patients must be admitted on a set day or within an admission time window. Multiple resources along the patients' clinical pathway are considered. The models which break down into a fixed and flexible admission date model are embedded in a rolling-horizon approach to evaluate two types of uncertainties (length of stay and non-availability of resources because of emergencies).

Other related works are Ceschia and Schaerf (2016) and Bilgin *et al.* (2012). The papers focus on the patient admission scheduling problem which assigns patients to hospital beds over a given time horizon. The objective is to maximize treatment efficiency, patient comfort and hospital utilization, while satisfying all necessary medical constraints and taking into consideration patient preferences as much as possible. Both papers are located on a more detailed level than we and Gartner and Kolisch (2014) do: Patients are assigned to beds and not to a capacity within a specialty. Another difference is the multi-criteria objective function while our goal is to maximize contribution margin.

More recent publications not included in Gartner and Padman (2017b) can be broken down into extensions of the patient admission scheduling (PAS) problem: Turhan and Bilgen (2017), for example, extend the PAS problem and develop a fix-and-relax method as a solution approach. Again, our model is different because of the objective function and constraints but also because we look into the patients' clinical pathway and therefore a mix of resources such as diagnostics, operating theatres and beds. Moreover, Turhan and Bilgen (2017) solve the model in a static way while we embed our model in a dynamic and rolling horizon approach.

The patient scheduling problem addressed by Bastos *et al.* (2018) and Burdett *et al.* (2017) maximizes the number of patients of each type that a hospital can accept which is also different from our problem because our model maximizes the contribution margin associated with the acceptance of patients and overtime costs. Another relevant paper is Burdett and Kozan (2018) who develop constructive algorithms and hybrid meta-heuristics to schedule clinical pathways. To conclude, Gartner and Kolisch (2014) is most relevant because of flexibility extensions in the admission, overtime, specialty and team assignment. These extensions are modeled using mathematical programming and incorporated in a rolling-horizon planning framework.

3. Problem description, model and examples

We will now present the parameters, decision variables, objective function and constraints of our mathematical model. It unifies and generalizes the two models proposed by Gartner and Kolisch (2014). Afterwards, examples will demonstrate the effectiveness of the flexibility extensions.

3.1. Parameters

Although there exist similarities in the parameters of Gartner and Kolisch (2014)'s problem description, we will now introduce all necessary sets, indices and constants.

3.1.1. Planning horizon, patients, activities, execution modes, LOS and CPs

Let $\mathcal{T} := \{1, 2, \ldots, T\}$ be the set of days with planning horizon T, let $\mathcal{S}$ denote the set of weeks (sennights) with S as last index and let $\mathcal{T}_s \subset \mathcal{T}$ denote the subset of days in week $s \in \mathcal{S}$. Elective patients are denoted by set $\mathcal{P}$, $\mathcal{A}$ denotes the set of all clinical activities to be scheduled and $\mathcal{A}_p \subset \mathcal{A}$ denotes the subset of activities for patient $p \in \mathcal{P}$. Let $\mathcal{M}_i$ denote the execution modes of an activity $i \in \mathcal{A}$. This will be used in two ways as the example in Section 3.5 will show: First, execution modes of an activity can be used to provide flexibility in assigning e.g. a surgical team to a surgery. Second, execution modes of an admission and discharge activity will be used to cope for multi-morbid patients that may be assigned to one out of multiple relevant specialties. Let $\mathcal{L}_p := \{0, 1, 2, \ldots, T\}$ denote the set of lengths of stay for patient $p \in \mathcal{P}$ in which the maximum length of stay cannot be longer than the planning horizon T. A patient who is not admitted has zero length of stay. 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 scheduled at least $d_{i,j}^{\min}$ days later than activity i. Given the graph $(\mathcal{A}, \mathcal{E})$, the admission and the discharge activity σ_p and ϕ_p, respectively for each patient $p \in \mathcal{P}$, we calculate for each activity $i \in \mathcal{A}$ the earliest day E_i and the latest day L_i on which the activity has to be scheduled with longest path methods (see, for example, Neumann *et al.* (2003)). Let $\mathcal{W}_i := \{E_i, E_i + 1, \ldots, L_i\}$ denote the time window of activity i. Once we have calculated the latest day L_{ϕ_p} in which the discharge activities can be scheduled, the index of the last week can be calculated by $S = \left\lceil \frac{\max_{p \in \mathcal{P}} L_{\phi_p}}{7} \right\rceil$. Accordingly, T can be set to the last day in the last week $T = 7 \cdot S$.

3.1.2. Hospital resources, capacity and demand

Scarce hospital resources $\mathcal{R}$ are depicted by three subsets. Day resources $\mathcal{R}^{\mathrm{d}} \subset \mathcal{R}$, human resources $\mathcal{R}^{\mathrm{h}} \subset \mathcal{R}$ and overnight resources $\mathcal{R}^{\mathrm{n}} \subset \mathcal{R}$. Each resource $k \in \mathcal{R}^{\mathrm{d}}$ has a day capacity $R_{k,t}^{\mathrm{day}}$ on day $t \in \mathcal{T}$, e.g. 1 slot available for trauma and orthopedics in the master surgical schedule on a Monday. Each human resource $k \in \mathcal{R}^{\mathrm{h}}$ has a capacity $R_{k,t}^{\mathrm{human}}$ on day $t \in \mathcal{T}$, e.g. 8 hours anesthesia support in the operating theater. In addition to the day-dependent capacity, there is a maximum week capacity of human resource $k \in \mathcal{R}^{\mathrm{h}}$ in week $s \in \mathcal{S}$ which we depict by $R_{k,s}^{\mathrm{week}}$, for example, 70 hours. The capacity demand of activity $i \in \mathcal{A}$ in mode $m \in \mathcal{M}_i$ on resource $k \in \mathcal{R}$ is $r_{i,m,k}$. An example that illustrates the concept of multiple execution modes is provided in Section 3.5.3. Capacity and demand are measured in hours for day and human resources. For overnight resources, we measure demand and capacity in beds.

3.1.3. Contribution margin, target working hours and overtime costs

We consider a healthcare system in which, given a patient $p \in \mathcal{P}$ and his length of stay $l \in \mathcal{L}_p$, the hospital receives a contribution margin $\pi_{p,l}$. This concept is applicable in particular to DRG systems with LOS and DRG-dependent contribution margin functions, see Gartner and Kolisch (2014) but applies to healthcare systems with other LOS and DRG-dependent contribution margin functions, too. Each human resource $k \in \mathcal{R}^{\text{h}}$ has a week target working time of B_k^{o}, e.g. a 40 hours week. If the target working time within a week is exceeded, the human resource can compensate that time with undertime in the following week. If overtime is not compensated in the following week, it is paid out and overtime costs c_k^{o} occur.

Table 1 provides an overview of all parameters and decision variables. The latter will be introduced next.

Table 1.: Sets, indices, constants and decision variables

Parameter	Description
$\mathcal{A}$	Set of activities
$\mathcal{A}_p$	Set of activities corresponding to patient $p \in \mathcal{P}$
α_p	Earliest admission date for patient $p \in \mathcal{P}$
B_k^{o}	Target working hours for human resource $k \in \mathcal{R}^{\text{h}}$ (e.g. 40 hours)
c_k^{o}	Costs per hour overtime
$d_{i,j}^{\min}$	Minimum time lag for precedence relation $(i, j) \in \mathcal{E}$
$\mathcal{E}$	Set of precedence relations
E_i	Earliest day to schedule activity $i \in \mathcal{A}$
$\mathcal{L}_p$	Set of possible lengths of stay for patient $p \in \mathcal{P}$
L_i	Latest day to schedule activity $i \in \mathcal{A}$
$\mathcal{M}_i$	Set of modes to schedule activity $i \in \mathcal{A}$
$\mathcal{P}$	Set of patients
ϕ_p	Discharge activity for patient $p \in \mathcal{P}$
$\pi_{p,l}$	Contribution margin of patient $p \in \mathcal{P}$ with a length of stay $l \in \mathcal{L}_p$
ϖ_p	1, if the surgery lead time equals 1 day, 0 otherwise
$\mathcal{R} := \mathcal{R}^{\text{d}} \bigcup \mathcal{R}^{\text{h}} \bigcup \mathcal{R}^{\text{n}}$	Set of resources (day, human and overnight, respectively)
ϱ_p	Recovery time of patient $p \in \mathcal{P}$
ρ_p	Recovery time induction activity of patient $p \in \mathcal{P}$
$r_{i,m,k}$	Capacity demand on resource $k \in \mathcal{R}$ if activity $i \in \mathcal{A}$ is scheduled in mode $m \in \mathcal{M}_i$
$R_{k,t}$	Maximum capacity of resource $k \in \mathcal{R}$ (e.g. 10 hours of a CT, number of available beds) on day $t \in \mathcal{T}$. The matrix unifies the day, human and overnight resource capacity matrices which are described next.
$R_{k,t}^{\text{day}}$	Maximum capacity of day resource $k \in \mathcal{R}^{\text{d}}$ (e.g. 1 slot in the master surgical schedule)
$R_{k,t}^{\text{human}}$	Maximum capacity of human resource $k \in \mathcal{R}^{\text{h}}$ (e.g. 8 hours)

$R^{\text{night}}_{k,t}$	Maximum capacity of overnight resource $k \in \mathcal{R}^{\text{n}}$ (e.g. 70 beds)
$R^{\text{week}}_{k,s}$	Maximum week capacity of human resource $k \in \mathcal{R}^{\text{h}}$ in week $s \in \mathcal{S}$ (e.g. 70 hours)
$R^{\text{real}}_{k,1}$	Realization of remaining resource capacity for resource $k \in \mathcal{R}$ at day $t = 1$ (e.g. 3.85 hours, 29 beds)
$\mathcal{S}$	Set of weeks (sennights)
S	Index of the last week within the set of weeks $\mathcal{S}$
S^{θ}_{i}	Vector of start times for activity i in rolling horizon iteration θ
σ_p	Admission activity for patient $p \in \mathcal{P}$
θ	Rolling horizon iteration
$\mathcal{T}$	Set of days
$\mathcal{T}_s \subset \mathcal{T}$	Subset of days in week $s \in \mathcal{S}$
$\mathcal{W}_i$	Set of consecutive days to schedule activity $i \in \mathcal{A}$

Decision variable	Description
$o_{k,s} \in \mathbb{R}_{\geq 0}$	Continuous variable to measure overtime for human resource $k \in \mathcal{R}^{\text{h}}$ in week $s \in \mathcal{S}$
$u^{\text{succ}}_{k,s} \in \mathbb{R}_{\geq 0}$	Continuous undertime variable for human resource $k \in \mathcal{R}^{\text{h}}$ that occurs in week $s+1$ w.r.t. week $s \in \mathcal{S}$
$x_{i,m,t}$	1, if activity i is scheduled in mode $m \in \mathcal{M}_i$ at day $t \in \mathcal{W}_i$, 0 otherwise
$y_{p,l}$	1, if patient $p \in \mathcal{P}$ is assigned to LOS $l \in \mathcal{L}_p$, 0 otherwise
z_p	1, if patient p is admitted, 0 otherwise

3.2. Decision variables

We extend the binary activity-to-day assignment variables used in Gartner and Kolisch (2014) to a mode dimension. We borrow the concept from multi-mode resource-constraint multi-project scheduling formulations (Wauters *et al.* (2016)). Accordingly,

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

Similar to Gartner and Kolisch (2014), we employ binary variables

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

which assign a patient to a LOS. If the target working time (usually 40 hours) is exceeded, overtime occurs, expressed by the real valued variable $o_{k,s} \in \mathbb{R}$ which counts the overtime for human resource $k \in \mathcal{R}^{\text{h}}$ in week $s \in \mathcal{S}$. Finally, whether a patient is admitted or not is indicated by the binary variables

$$z_p = \begin{cases} 1, & \text{if patient } p \in \mathcal{P} \text{ is admitted} \\ 0, & \text{otherwise.} \end{cases}$$

3.3. Objective function

Having introduced all necessary parameters and decision variables, the objective function reads as given by Equation (1).

Maximize $z =$

$$\sum_{p\in\mathcal{P}}\sum_{l\in\mathcal{L}_p} \pi_{p,l}\cdot y_{p,l} - \sum_{k\in\mathcal{R}^{\mathrm{h}}}\sum_{s\in\mathcal{S}\setminus S} c^{\mathrm{o}}_k \cdot \left(o_{k,s} - \left(B^{\mathrm{o}}_k - \sum_{i\in\mathcal{A}}\sum_{m\in\mathcal{M}_i}\sum_{t\in\mathcal{T}_{s+1}\cap\mathcal{W}_i} r_{i,m,k}\cdot x_{i,m,t}\right)^+\right)^+ \tag{1}$$

The first term in the objective function denotes the contribution margin for all patients while the second term denotes overtime costs. Herein, the term $\left(B^{\mathrm{o}}_k - \sum_{i\in\mathcal{A}}\sum_{m\in\mathcal{M}_i}\sum_{t\in\mathcal{T}_{s+1}\cap\mathcal{W}_i} r_{i,m,k}\cdot x_{i,m,t}\right)^+$ calculates the undertime that occurs in week $s+1$ which is the difference between the target working time and human resource requirements of all activities in that week while $(x)^+$ denotes $\max\{0,x\}$. The net overtime of week s can then be expressed as the difference of overtime in week s, $o_{k,s}$, and undertime in week $s+1$. Multiplying (net overtime)$^+$ of human resource k with the overtime cost per hour c^{o}_k and summing up over all human resources give the total overtime costs. Note that this objective function is non-linear. A linearization is provided in Appendix B.

3.4. Constraints

In what follows, we add constraints to our model which we break down by clinical pathways, day and overnight resource constraints and working time regulations.

3.4.1. Clinical pathways, day and overnight resource constraints

Constraints (2) ensure minimum time lags between all consecutive activities of an admitted patient.

$$\sum_{m\in\mathcal{M}_j}\sum_{t\in\mathcal{W}_j} t\cdot x_{j,m,t} \geq \sum_{m\in\mathcal{M}_i}\sum_{t\in\mathcal{W}_i} t\cdot x_{i,m,t} + d^{\min}_{i,j}\cdot z_p \quad \forall p\in\mathcal{P}, (i,j)\in\mathcal{E}_p \tag{2}$$

Day resource constraints (3) ensure that the demand for each day resource does not exceed its capacity.

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

Human resource constraints (4) ensure that the demand for each human resource does not exceed its day-dependent capacity.

$$\sum_{i \in \mathcal{A}: t \in \mathcal{W}_i} \sum_{m \in \mathcal{M}_i} r_{i,m,k} \cdot x_{i,m,t} \leq R_{k,t}^{\text{human}} \qquad \forall k \in \mathcal{R}^{\text{h}}, t \in \mathcal{T} \tag{4}$$

Overnight resource constraints (5) guarantee that admitted patients do not exceed bed capacity for each specialty. An example is provided in Section 3.5.2.

$$\sum_{p \in \mathcal{P}} \Bigg(\sum_{\substack{m \in \mathcal{M}_{\sigma_p}: \\ r_{\sigma_p,m,k}=1}} \sum_{\tau=E_{\sigma_p}}^{\min\{t, L_{\sigma_p}\}} x_{\sigma_p,m,\tau} - \sum_{\substack{m \in \mathcal{M}_{\phi_p}: \\ r_{\phi_p,m,k}=1}} \sum_{\tau=E_{\phi_p}}^{\min\{t, L_{\phi_p}\}} x_{\phi_p,m,\tau} \Bigg) \leq R_{k,t}^{\text{night}} \quad \forall k \in \mathcal{R}^{\text{n}}, t \in \mathcal{T} \tag{5}$$

Constraints (6) ensure that once a patient is admitted to a specialty which is represented by the admission activity's mode, he must be discharged from that specialty.

$$\sum_{t \in \mathcal{W}_{\sigma_p}} x_{\sigma_p,m,t} = \sum_{t \in \mathcal{W}_{\phi_p}} x_{\phi_p,m,t} \qquad \forall p \in \mathcal{P}, m \in \mathcal{M}_{\sigma_p} \tag{6}$$

3.4.2. Working time regulations

Constraints (7) link the overtime variables $o_{k,s}$ for human resource $k \in \mathcal{R}^{\text{h}}$ and week $s \in \mathcal{S}$ with the capacity requirement and the target working time.

$$B_k^{\text{o}} \geq \sum_{i \in \mathcal{A}} \sum_{t \in \mathcal{T}_s: t \in \mathcal{W}_i} \sum_{m \in \mathcal{M}_i} r_{i,m,k} \cdot x_{i,m,t} - o_{k,s} \qquad \forall k \in \mathcal{R}^{\text{h}}, s \in \mathcal{S} \tag{7}$$

3.4.3. LOS calculation and assignment

Constraints (8) set the length of stay as the difference between the day of admission and the day of discharge for each patient.

$$\sum_{l \in \mathcal{L}_p} l \cdot y_{p,l} = \sum_{m \in \mathcal{M}_{\phi_p}} \sum_{t \in \mathcal{W}_{\phi_p}} t \cdot x_{\phi_p,m,t} - \sum_{m \in \mathcal{M}_{\sigma_p}} \sum_{t \in \mathcal{W}_{\sigma_p}} t \cdot x_{\sigma_p,m,t} \qquad \forall p \in \mathcal{P} \tag{8}$$

Constraints (9) guarantee that exactly one length of stay is assigned to each patient. If a patient is not admitted, his LOS becomes zero.

$$\sum_{l \in \mathcal{L}_p} y_{p,l} = 1 \qquad \forall p \in \mathcal{P} \tag{9}$$

3.4.4. Activity-to-mode-and-day assignment and admission decision

Constraints (10) ensure that each activity is scheduled at most once which implies that each activity which is processed, exactly one processing mode is chosen.

$$\sum_{m \in \mathcal{M}_i} \sum_{t \in \mathcal{W}_i} x_{i,m,t} \leq 1 \qquad \forall i \in \mathcal{A} \tag{10}$$

Constraints (11) ensure that once a patient is admitted, all his activities are scheduled. Otherwise, none of his activities are scheduled.

$$\sum_{i \in \mathcal{A}_p} \sum_{t \in \mathcal{W}_i} \sum_{m \in \mathcal{M}_i} x_{i,m,t} = |\mathcal{A}_i| \cdot z_p \qquad \forall p \in \mathcal{P} \tag{11}$$

3.4.5. Decision variables

Variable definitions and their domains are given by (12)–(15).

$$0 \leq o_{k,s} \leq R^{\text{week}}_{k,s} - B^{\text{o}}_k \qquad \forall k \in \mathcal{R}^{\text{h}}, s \in \mathcal{S} \tag{12}$$

$$x_{i,m,t} \in \{0,1\} \qquad \forall i \in \mathcal{A}, m \in \mathcal{M}_i, t \in \mathcal{W}_i \tag{13}$$

$$y_{p,l} \in \{0,1\} \qquad \forall p \in \mathcal{P}, l \in \mathcal{L}_p \tag{14}$$

$$z_p \in \{0,1\} \qquad \forall p \in \mathcal{P} \tag{15}$$

3.5. Examples

In what follows, we give examples of overtime reduction, bed allocation and show a sample test instance that we solve to optimality.

3.5.1. Overtime reduction examples

We now provide three examples to demonstrate cases where overtime cannot be reduced, overtime is partially reduced and where overtime can be reduced entirely. All of the cases consider the reduction of overtime in week $s+1$, given overtime that occurred in week s. To give an example, we assume one human resource, denoted by $k = 1$ which is a nurse who has worked 50 hours in week $s = 1$ and builds up ten hours of overtime ($o_{1,1} = 10$), based on $B^{\text{o}}_1 = 40$ hours per week target working time as given by her contract. The result when the nurse works 50, 35 and 0 hours in week $s = 2$ are shown in Table 2. As a consequence, overtime cannot be reduced, is partially reduced and completely reduced, respectively.

Table 2. Overtime reduction examples

Case	Overtime in week $s = 1$	Undertime in week $s = 2$	Remaining overtime in week $s = 1$
No reduction	10	$(40 - 50)$	10
Partial reduction	10	$(40 - 35)$	5
Complete reduction	10	$(40 - 0)$	0

As can be seen, if the nurse works 50 hours in week $s = 2$, the nurse cannot reduce overtime that was built up in week $s = 1$. As a consequence, the second part of objective function (1) reads as $(10 - (40 - 50)^+)^+ = 10$ and overtime costs c_1^{o} are multiplied by 10.

In the case of partial overtime reduction, we assume that the nurse works 35 hours in week $s = 2$. As a result, the second part of Objective function (1) reads as $(10 - (40 - 35)^+)^+ = 5$ and overtime costs c_1^{o} are multiplied by 5.

In the case of complete overtime reduction, we assume that the nurse has week $s = 2$ overtime compensation through time off. As a result, the second part of objective function (1) reads as $(10 - (40 - 0)^+)^+ = 0$. This means that no overtime costs are paid out.

3.5.2. A flexible specialty allocation example

In what follows, we provide an example to demonstrate the concept of flexible specialty allocation with the following parameters: We have a planning horizon of $\mathcal{T} := \{1, \ldots, 7\}$ days, overnight resources $\mathcal{R}^{\mathrm{n}} := \{1, 2\}$ and a single patient $\mathcal{P} := \{1\}$ with admission time window $\mathcal{W}_{\sigma_1} = \{1, 2, 3\}$ and discharge time window $\mathcal{W}_{\phi_1} = \{5, 6, 7\}$. We assume that the admission activity σ_1 and discharge activity ϕ_1 can be executed in two modes $\mathcal{M}_{\sigma_1} = \{1, 2\}$ and $\mathcal{M}_{\phi_1} = \{1, 2\}$. This means that there is the option to admit the patient to the surgical or to the internal medicine specialty as we assume that beds are not ring-fenced. The mode-dependent specialty requirements are thus $r_{\sigma_1,1,1} = 1$, $r_{\sigma_1,1,2} = 0$, $r_{\sigma_1,2,1} = 0$, and $r_{\sigma_1,2,2} = 1$. Table 3 provides an example specialty allocation. The first four rows provide the decision variables $x_{\sigma_1,m,t}$ and $x_{\phi_1,m,t}$, respectively. We assume that the x-variables in the solution come up with the following assignment: $x_{\sigma_1,1,2} = 1$ and $x_{\phi_1,1,7} = 1$ and 0 otherwise. This clearly satisfies constraints (10) because $\sum_{m \in \{1,2\}} \sum_{t \in \mathcal{W}_i} x_{i,m,t} \leq 1 \quad \forall i \in \{\sigma_1, \phi_1\}$. Now, the last two rows of the table show how the overnight resources are allocated. As can be seen, the patient requires a bed on overnight resource $k = 1$ starting with the night between days 2 and 3 until the night between days 6 and 7. The table also reveals that overnight resource $k = 2$ is not allocated as shown in the last row.

3.5.3. An example for the flexible hospital-wide elective patient flow problem

In what follows we give an example with four patients of which the LOS dependent contribution margin is given in Table 4.
Consider the set of weeks $\mathcal{S} := \{1, 2\}$ and the set of days $\mathcal{T} := \{1, \ldots, 14\}$. Activities $i \in \mathcal{A}$, resources $\mathcal{R}^{\mathrm{d}} := \{1, 2, \ldots, 9\}$, $\mathcal{R}^{\mathrm{h}} := \{10, 11, 12, 13\}$, and $\mathcal{R}^{\mathrm{n}} := \{14, 15\}$ are shown in Table 5.

Table 3. A bed allocation example

$t \in \mathcal{T}$	1	2	3	4	5	6	7
$x_{\sigma_1,1,t}$	0	1	0	–	–	–	–
$x_{\phi_1,1,t}$	–	–	–	–	0	0	1
$x_{\sigma_1,2,t}$	0	0	0	–	–	–	–
$x_{\phi_1,2,t}$	–	–	–	–	0	0	0
$k=1: \quad \sum_{p \in \mathcal{P}} \left(\sum_{m \in \mathcal{M}_\sigma} \sum_{\tau=E_{\sigma_1}}^{\min\{t, L_{\sigma_1}\}} x_{\sigma_1,m,\tau} - \sum_{m \in \mathcal{M}_\phi} \sum_{\tau=E_{\phi_1}}^{\min\{t, L_{\phi_1}\}} x_{\phi_1,m,\tau} \right)$	0	1	1	1	1	1	0
$k=2: \quad \sum_{p \in \mathcal{P}} \left(\sum_{m \in \mathcal{M}_\sigma} \sum_{\tau=E_{\sigma_1}}^{\min\{t, L_{\sigma_1}\}} x_{\sigma_1,m,\tau} - \sum_{m \in \mathcal{M}_\phi} \sum_{\tau=E_{\phi_1}}^{\min\{t, L_{\phi_1}\}} x_{\phi_1,m,\tau} \right)$	0	0	0	0	0	0	0

Table 4. LOS dependent contribution margins (in Euros) for the four patients in the example

t	5	6	7	12	13	14
$\pi_{1,t}$	3,772.67	3,711.80	3,650.94	-	-	-
$\pi_{2,t}$	3,498.41	3,436.15	3,373.90	-	-	-
$\pi_{3,t}$	-	-	-	3,292.69	3,204.17	3,161.10
$\pi_{4,t}$	-	-	-	3,833.47	3,797.35	3,763.30

Table 5. Activities (a), day, human and overnight resources (b) and the cyclic MSS (c)

$i \in \mathcal{A}$	Description
1	Admission of pat. 1
2	Spine CT for pat. 1
3	Spinal surgery of pat. 1
4	Discharge of pat. 1
5	Admission of pat. 2
6	Arteriography for pat. 2
7	Stent implantation of pat. 2
8	Discharge of pat. 2
9	Admission of pat. 3
10	X-ray for pat. 3
11	Herniotomy of pat. 3
12	Discharge of pat. 3
13	Admission of pat. 4
14	Sonography for pat. 4
15	Cholecystectomy of pat. 4
16	Discharge of pat. 4

(a)

$k \in \mathcal{R}$	Description	$R_{k,t}$
1	Radiology unit	0.5h on workdays, 0 otherwise
2, . . . , 5	MSS slot 1, . . . , 4 for surgical specialty	See cyclic MSS
6, . . . , 9	MSS slot 1, . . . , 4 for internal medicine specialty	See cyclic MSS
10	Surgeon 1	2h Mo.–Fri., 0 otherwise
11	Surgeon 2	2h Mo.–Fri., 0 otherwise
12	Surgical nurse 1	2h Mo.–Fri., 0 otherwise
13	Surgical nurse 2	2h Mo.–Fri., 0 otherwise
14	Surgical ward	2 beds on workdays, 1 otherwise
15	Internal medicine ward	2 beds on workdays, 1 otherwise

(b)

	1	2	3	4	5	6	7
Slot 1 (8–10 a.m.)	Surgical	Int. Med.	–	–	–	–	–
Slot 2 (10–12 a.m.)	–	–	Surgical	Int. Med.	–	–	–
Slot 3 (12–14 a.m.)	–	Surgical	Int. Med.	–	–	–	–
Slot 4 (14–16 a.m.)	–	–	Int. Med.	–	Surgical	–	–

(c)

Clinical pathways and resource requirements for day resources are shown in Figure 1. Each of the surgical activities $i \in \{3, 7, 11, 15\}$ have 4 modes. Activity $i = 3$ belongs to the surgical specialty and therefore any of the 4 surgical specialty slots can be assigned. In contrast, activity $i = 7$ is a stent implantation and therefore, the patient belongs to the internal medicine specialty. Since the internal medicine specialty has 4 slots (see Table 5(c)), there are 4 options or assignment modes. The same principle is applied to activities 11 and 15: because they can be executed either in one of the 4 surgical MSS slots or in one of the 4 internal medicine MSS slots. All other activities have, for simplicity 1 mode. The generation of the surgery modes has to be done by preprocessing where the number of modes depends on the available surgery slots.

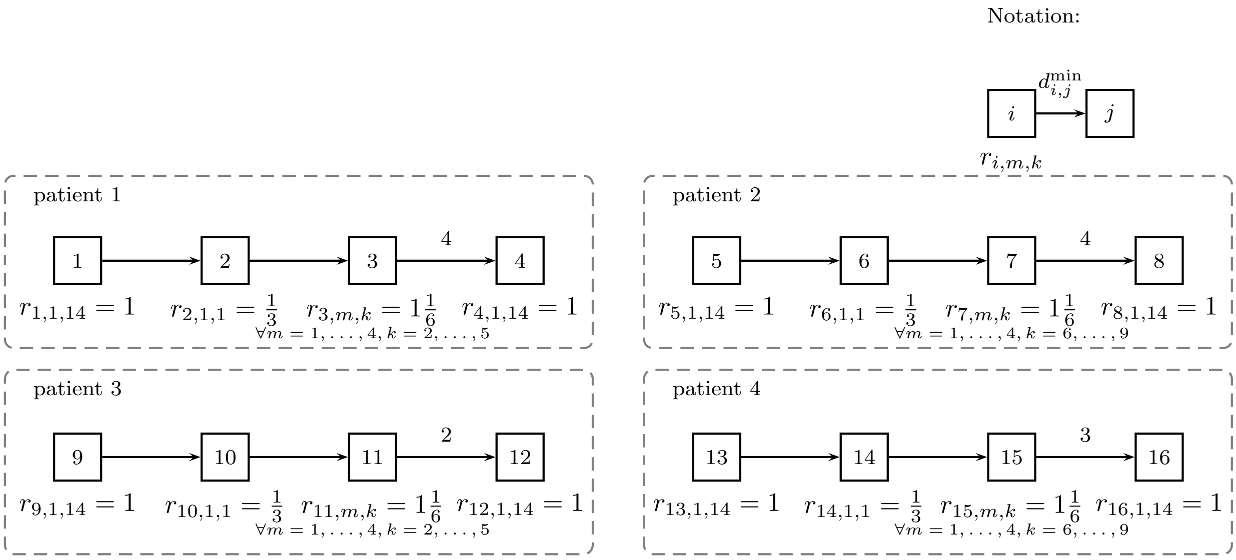


Figure 1. Clinical pathways for the four patients

In Figure 1, $r_{1,1,14} = 1$ denotes that activity 1 (the admission activity for patient 1) requires 1 bed from the surgical ward. The minimum time lag $d_{3,4}^{\min} = 4$ between activity 3 and 4 in Figure 1 denotes that at least four days of recovery time have to pass between the surgery and the discharge of patient 1.

Table 6(a) provides earliest and latest days to schedule the activities as obtained by longest path calculation. Overtime costs c_k^o per hour overtime for the two types of human resources are given in Table 6(b).

Table 6. Earliest and Latest starts, and Number of modes of each activity (a) and Overtime costs per hour in Euros (b)

$i \in \mathcal{A}$	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
E_i	1	1	1	5	1	1	1	5	8	8	8	12	8	8	8	12
L_i	3	3	3	7	3	3	3	7	11	11	11	14	11	11	11	14
$\lvert\mathcal{M}_i\rvert$	2	1	4	2	2	1	4	2	2	1	4	2	2	1	4	2

(a)

$k \in \mathcal{R}^{\mathrm{h}}$	c^{o}_k
1, 2	170
3, 4	100

(b)

Table 7 provides a summary of the parameters and the number of decision variables (DV) and constraints (Cts.) for model (1)–(15).

Table 7. Summary statistics of the sample instance

$\lvert\mathcal{T}\rvert$	$\lvert\mathcal{P}\rvert$	$\lvert\mathcal{A}\rvert$	$\sum_{i\in\mathcal{A}} \lvert\mathcal{W}_i\rvert$	$\lvert\mathcal{E}\rvert$	$\lvert\mathcal{R}^{\mathrm{d}}\rvert$	$\lvert\mathcal{R}^{\mathrm{n}}\rvert$	$\lvert\mathcal{R}^{\mathrm{h}}\rvert$	$\lvert\mathcal{L}_p\rvert$	$\lvert\mathcal{S}\rvert$	$\sum_{i\in\mathcal{A}} \lvert\mathcal{M}_i\rvert$	#DV	#Cts.
14	4	16	64	12	9	2	4	8	2	36	130	262

The total contribution margin when solving the test instance is 14,127.24 Euros and all four patients are admitted. However, if no overtime is allowed by fixing variables $c^{\mathrm{o}}_{k,s} = 0 \quad \forall k \in \mathcal{R}^{\mathrm{h}}, s \in \mathcal{S}$, then only three patients are admitted, leading to a contribution margin of 10,501.51 Euros.

4. Rolling horizon planning

It is common practice that hospitals undertake daily reallocation decisions for single resources including the operating theater and specialties. We therefore embed our model in a rolling horizon approach in which each day, a problem instance is solved. This approach has been tested successfully by Gartner and Kolisch (2014). Their approach can be summarized as follows. In each run, the most recent data is incorporated as parameters into a mathematical model. Deterministic information is assumed to be available at the day of planning, including realizations of recovery time and remaining resource capacities. In this study, we differ from their approach because we employ expected values for remaining recovery times and remaining resource capacities of future days that follow *lognormal* distributions conditioned on DRG and resource type, respectively. We have chosen this based on related work (see Min and Yih (2010) and Gartner and Padman (2016)) as well as results from the fitting of our data to lognormal distributions, see Appendix C. Because there is no set capacity for emergency patients, we have to compute with the remaining resource capacity available for scheduling our set of elective patients.

4.1. General framework for the rolling horizon planning

Each iteration of the rolling horizon planning is denoted by θ, initialized to $\theta := 0$ at the beginning, and incremented after each iteration by $\theta \leftarrow \theta + 1$. Let S_i^θ be the vector of start times for activities $i \in \mathcal{A}$ obtained by solving the instances for each rolling horizon iteration θ. The activity that induces recovery for patient $p \in \mathcal{P}$ is denoted by ρ_p. Each subset $\mathcal{A}_p \subset \mathcal{A}$ contains an ordered tuple $(1, \ldots, \rho_p, \ldots, \phi_p)$ of activities that belong to patient $p \in \mathcal{P}$. Expected recovery times of patient $p \in \mathcal{P}$ are denoted by $E[\varrho_p]$. The expected remaining capacity of resource $k \in \mathcal{R}$ on day $t \in \mathcal{T}$ is denoted by $E[R_{k,t}]$. The realization of remaining capacity available for elective patients is denoted by $R_{k,t}^{\text{real}}$. More precisely, it is assumed that complete information about the realized capacity is available only for the beginning of the rolling horizon (day $t = 1$). Thus, $R_{k,1}^{\text{real}}$ is considered instead of $E[R_{k,1}]$. Next, we have the surgery lead time ϖ_p which is 1, if patient $p \in \mathcal{P}$ has to wait at least one day from the admission until the surgery date. Otherwise, $\varpi_p = 0$. This means that the patient can have surgery at the beginning of the rolling horizon planning. Next, arriving patient requests are generated and all activities and further parameters such as precedence relations are adapted. At the same step, the remaining resource capacity is set for the first day of the rolling horizon planning $t = 1$. Afterwards, time windows (amongst remaining recovery times), precedence relations and time lags are recalculated. Then, the instance is solved and the schedule is stored for each day θ in vectors S_i^θ.

4.2. Time window and recovery time updating procedure

The procedure to adapt the time windows during each iteration θ follows Figure 2. We introduce α_p which is the earliest admission date for patient $p \in \mathcal{P}$. This can be in the future and may become fixed based on previous rolling planning steps. It can also be at the beginning of the rolling horizon planning or in the past. Now, to adapt the time windows, we check in the first step whether or not patients have been discharged in earlier iterations of the rolling horizon planning. If the discharge activity ϕ_p of patient $p \in \mathcal{P}$ has been scheduled, we set each time window $\mathcal{W}_i$ for each activity $i \in \mathcal{A}_p$ empty. This prevents that any activity of the corresponding patient is scheduled which implies that no decision variables for that particular patient are handed over to the solver. Therefore, those activities will no longer be scheduled. Second, we check for day θ, whether the patient has been admitted or is scheduled for admission in future days. If patient $p \in \mathcal{P}$ is admitted in the current day or has already been admitted but not discharged, we first check, whether recovery induction activity ρ_p has been scheduled. If it has not yet been undertaken, the earliest start dates of all activities of the pre-recovery stage of patient $p \in \mathcal{P}$ are set to 1. Activities $\rho_p \leq i < \phi_p$ cannot start before $1 + \varpi_p$ because of a potential surgery preparation time after the admission of patient $p \in \mathcal{P}$. Accordingly, the earliest start dates are set to 1 for all pre-surgical activities and $1 + E[\rho_p]$ for the discharge activity. Second, if the patient is in the recovery stage, the earliest start dates of all activities except the discharge activity are set to day 1 while the earliest discharge date is obtained by the random number obtained from the lognormal distribution minus the recovery time observed so far. Note that the recovery stage may be over but not all activities (e.g. post surgical activities) have been executed for patient $p \in \mathcal{P}$ yet. Therefore, we set the latest start of all other activities to the latest day of the planning horizon T. Finally, we ensure that the time windows of all activities that have been scheduled in former days of the rolling horizon are empty.

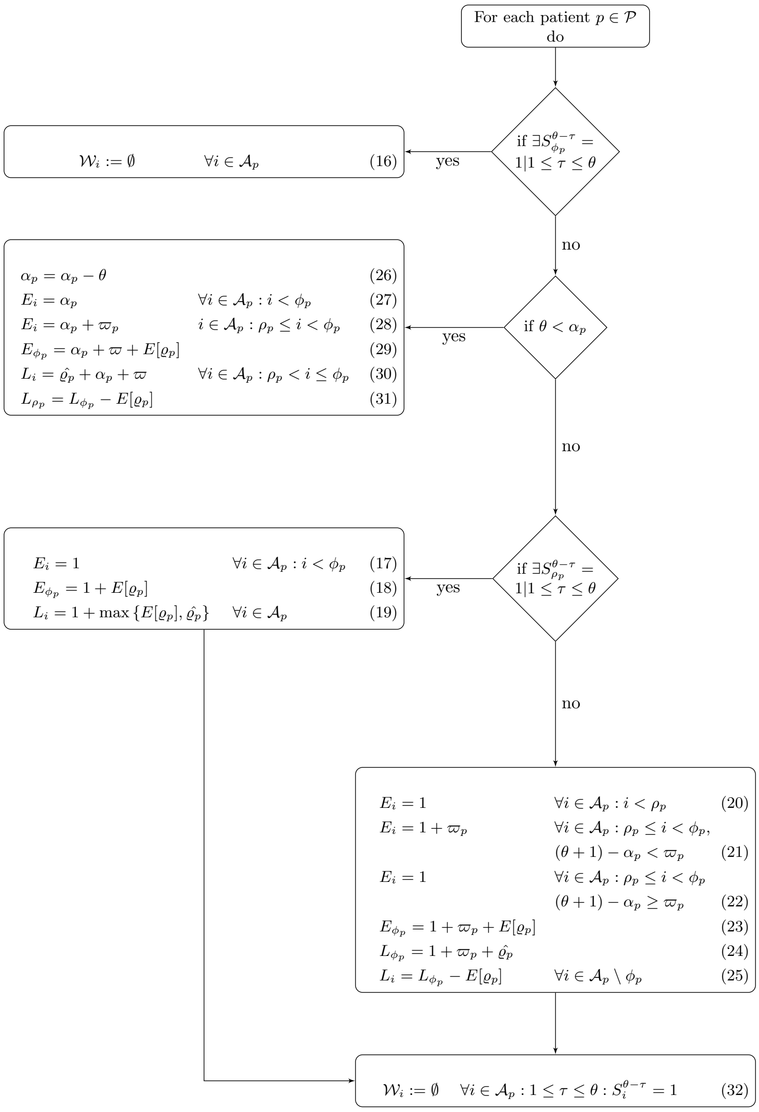


Figure 2. Time window adaptation sub-routine (Gartner (2015))

5. Experimental investigation

5.1. Data, test instances and evaluation measures

We employed data from a collaborating hospital in the vicinity of Munich, Germany. To test our model, we created four test instances by varying the number of patients from 45, 60, 75 and 90 patients labeled by test instance 1, 2, 3 and 4 respectively. We compare the rolling horizon planning and the perfect information solution on three metrics while we switch the allowance for overtime on and off. In the perfect information solution, we assume that the recovery times and the remaining resource capacities are known. The results are broken down by an economic analysis, waiting time of elective patients seeking admission, and average lengths of stay, reported in Sections 5.2, 5.3 and 5.4, respectively.

We generated data based on Gartner and Kolisch (2014)'s test instances. The major difference is, however, that the admission decision and the assignment of patients to specialties is flexible. For the admission decision and the admission time window, we chose α_p which is the patient's actual admission date and created a 7-day admission time window around it. Furthermore, we looked into each of the patient's co-morbidities to decide whether or not they are eligible for a flexible specialty assignment. The resource capacity of human resources was split into two anesthesia teams, two surgical teams and two operating theatre nursing teams with overtime costs set to the German regulations for hospital staff. As our approach takes into account lognormal distributed demand and capacity, we used the parameters as estimated from the statistical fit (see Appendix C). Across all rolling horizon iterations, the test instances had, on average, 303,321 decision variables and 6,716 constraints which required, on average a solution time of 3.9s This is an increase in model complexity and an (acceptable) increase in solution time as compared to Gartner and Kolisch (2014) where the average computation time was 0.5s.

5.2. Economic analysis

We run the rolling horizon planning as well as the model using perfect information and compare the scenarios with and without overtime. Figure 3 shows the results of our economic analysis.

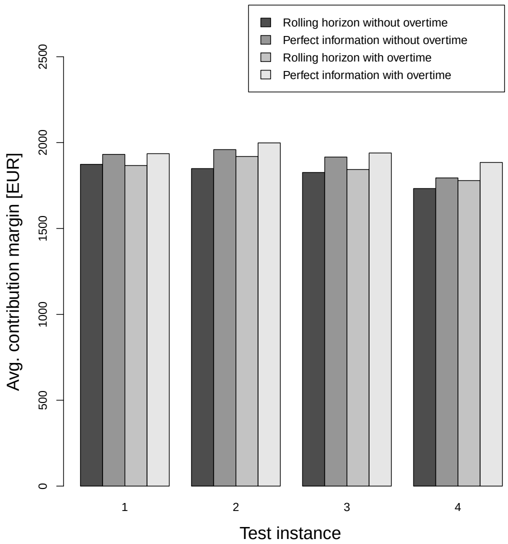


Figure 3. Rolling horizon vs. perfect information

The figure reveals differences between the perfect information solution and the rolling horizon planning which are between 3.0-5.6%. Another observation is that for the rolling horizon planning, the average contribution margin per patient decreases with increasing index of the test instances. One explanation for this phenomenon is that the hospital tries to admit patients in order to increase total contribution margin. However, resources become more and more scarce such that LOS increases, see Section 5.4, costs increase and as a consequence, average contribution margin per patient decreases. A closer investigation of overtime costs reveals that no overtime costs were generated at the end of the rolling horizon planning.

5.3. Waiting time analysis

Figure 4 shows the average waiting time for admitted elective patients.

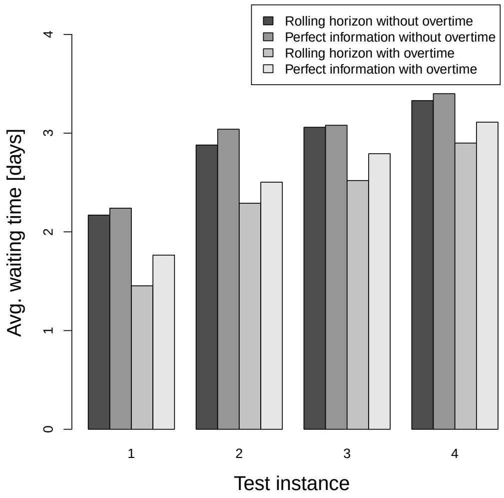


Figure 4. Rolling horizon vs. perfect information

The figure reveals that an increasing number of patients seeking for admission increases the waiting time. As can be seen, the difference between the average waiting time obtained by the rolling horizon approach and the one obtained by using perfect information is relatively small in the case when no overtime is allowed (0.9-5.7%). The differences are higher when overtime is allowed and vary between 4.5-21.4%. A more detailed analysis reveals that waiting times are below 4 days and acceptable for non-urgent elective patients.

5.4. Length of stay analysis

Figure 5 shows the average lengths of stay for the different setups.

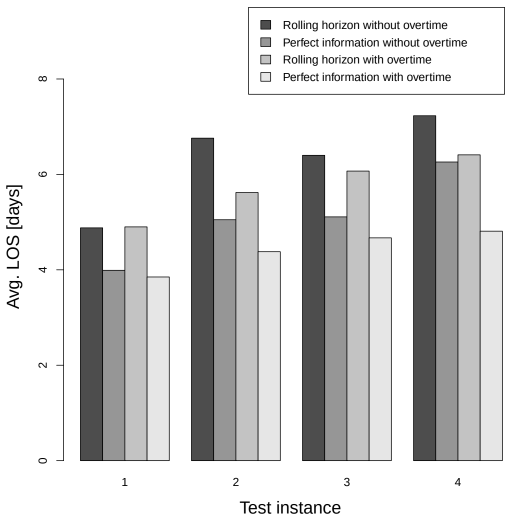


Figure 5. Rolling horizon vs. perfect information

We observe that the length of stay in the setting with perfect information is lower as compared to the rolling horizon planning. The difference is between 15.6-33.7%, depending on the instances. For the rolling horizon planning with overtime we observe a strictly monotonic increase of LOS along with increasing number of patients seeking admission. One explanation for this is, as explained earlier, that the model tries to admit patients to increase contribution margin while LOS can increase because resource capacity becomes more and more scarce.

5.5. Experiments to evaluate resource capacity variation

When schedulers assign elective patients to scarce hospital resources, they are faced with the problem of uncertain emergency patient demand. Since this demand can vary, variance in remaining resource capacity for elective patients should be taken into account. To this end, we ran experiments from no variance to a variance of 1.0 in the lognormal distributed resource capacity that can be allocated by elective patients. The results are shown in Figure 6. The figure shows that a higher variance in remaining resource capacity leads to a substantial decrease in contribution margin. A closer analysis of the results revealed that there is a 9.11% gap between the rolling horizon planning and the perfect information contribution margin for a variance of 1.0. In this case, the length of stay is substantially longer as compared to the case of the low variance scenarios. As a consequence, costs are increased because of the increased LOS and therefore contribution margins are lower.

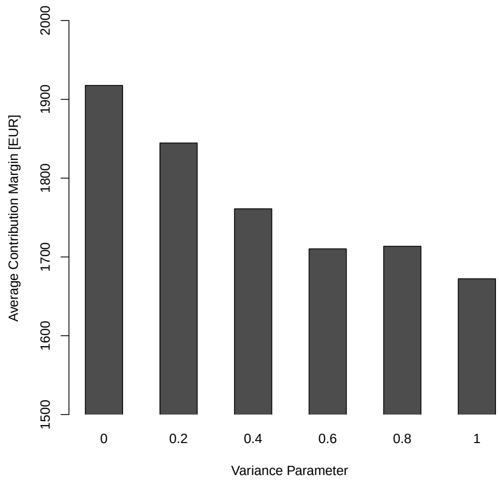


Figure 6. Average contribution margin per patient for different levels of variance in the resource capacities available for elective patients

As a conclusion of the experimental investigation, the benefits of the additional features of our model as compared to Gartner and Kolisch (2014) are as follows: The economic analysis has revealed that the model may use overtime temporarily to cope with new admissions or to ensure patients get discharged without delays. At the end of the planning horizon, however, it would not be paid out. The waiting time analysis showed that the model may prefer to substantially delay the admission of patients to maximize contribution margin. This is a benefit from our variable admission decisions.

Finally, the resource capacity variation experiments have allowed us to evaluate the impact of variation on contribution margins. We achieved this by varying the variance parameter of the lognormal-distributed remaining resource capacities.

6. Discussion

The hospital under consideration does not make transfer decisions across specialties at the point when patients are scheduled for admission. This means that, before admission, a multi-morbid patient would not be scheduled on an internal medicine ward for two days, followed by a recovery on a surgical specialty ward for the remainder of her hospital stay. Also, during the patient's hospital stay (after admission) it would prolong LOS if a patient was scheduled on one specialty and then transferred to another specialty with a different care team. This is different to assumptions made in bed assignment papers such as Demeester *et al.* (2009) where, within a specialty, bed re-assignments are much more common.

In the rolling horizon approach, we use the lognormal distribution to approximate the patients' length of stay and remaining resource capacity. This is different from the approach from Gartner and Kolisch (2014) who, for each patient in the system, used each patients' empirical length of stay distribution. The assumption of using lognormal distribution was based on the statistical test and the literature that supported it, see Min and Yih (2010).

7. Conclusion

In this paper we have presented a discrete optimization model for the problem of scheduling elective patients hospital-wide with a combined objective that maximizes DRG and LOS-dependent contribution margin and minimizes overtime costs. Our experimental analysis revealed that, if overtime flexibility is allowed, up to 5% increase in contribution margin can be achieved by reducing length of stay by up to 30%. At the same time, allowing for overtime can reduce access times by up to 33%.

Limitations in the modelling, computational study and usability in practice may include that the model does not account for tasks that clinicians have to do besides the treatment of patients. This workload can come from watch-lists, ward rounds or pre-medication services. However, these tasks can be added as additional jobs to be scheduled by the model. Alternatively, in the case of pre-medication ambulance workload, clinicians may have a reduced resource capacity on the given day. A drawback of the computational study is that the rolling horizon planning approach doesn't guarantee optimality and is only used to plan from one day to another, rather than looking at multiple stages of the admission and scheduling process.

Future work will incorporate other duties of staff into the model including the scheduling of individuals' tasks. Another extension is to embed the approach into Gartner and Padman (2017a)'s E-HOSPITAL platform to train managers how to schedule elective patients flexibly, effectively and efficiently in their day to day work. Another extension is the formulation of the problem as a multi-stage stochastic program.

Acknowledgement

The authors sincerely thank the associate editor and the four anonymous referees for their careful review and excellent suggestions for improvement of this paper.

Appendix A. Abbreviations

Table A1.: List of abbreviations

Abbreviation	Description
Avg.	Average
CP	Clinical pathway
Cts.	Constraints
DRG	Diagnosis-related group
DV	Decision variables
GUI	Graphical user interface
ICU	Intensive care unit
MSS	Master surgical schedule
OR	Operating room
PAS	Patient Admission Scheduling

Appendix B. Linearization of the objective function

Model formulation (1)–(15) is non-linear and can be linearized by introducing continuous undertime variables $u_{k,s}^{\text{succ}} \in \mathbb{R}_{\geq 0}$ that replace $\left(B_k^{\text{o}} - \sum\limits_{i \in \mathcal{A}} \sum\limits_{m \in \mathcal{M}_i} \sum\limits_{t \in \mathcal{T}_{s+1} \cap \mathcal{W}_i} r_{i,m,k} \cdot x_{i,m,t}\right)^+$ in the objective function. The variables denote the undertime for a human resource $k \in \mathcal{R}^{\text{h}}$, that occurs in the successor week $s+1$, based on week $s \in \mathcal{S} \setminus |\mathcal{S}|$. This substitution results in $\left(o_{k,s} - u_{k,s}^{\text{succ}}\right)^+$, which is replaced now by another continuous variable $o_{k,s}^{\text{net}} \in \mathbb{R}_{\geq 0}$. This represents the overtime, which cannot be recompensed by the under-time in week $s+1$. This linearization leads to the following additional constraints:

$$u_{k,s}^{\text{succ}} \geq \left(B_k^{\text{o}} - \sum_{i \in \mathcal{A}} \sum_{m \in \mathcal{M}_i} \sum_{t \in \mathcal{T}_{s+1} \cap \mathcal{W}_i} r_{i,m,k} \cdot x_{i,m,t}\right) \quad \forall k \in \mathcal{R}^h, s \in \mathcal{S} \setminus |\mathcal{S}| \tag{B1}$$

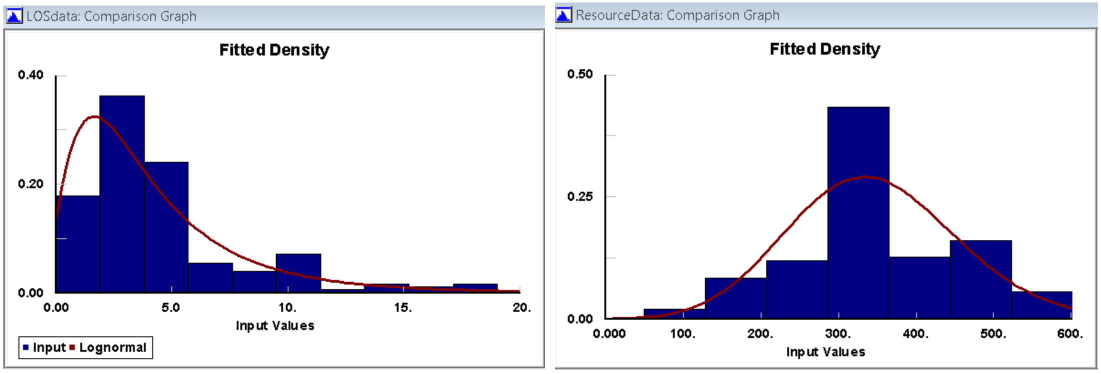
$$o_{k,s}^{\text{net}} \geq \left(o_{k,s} - u_{k,s}^{\text{succ}}\right) \quad \forall k \in \mathcal{R}^h, s \in \mathcal{S} \tag{B2}$$

As a result, our objective function can be simplified to:

$$z = \sum_{p \in \mathcal{P}} \sum_{l \in \mathcal{L}_p} \pi_{p,l} \cdot y_{p,l} - \sum_{k \in \mathcal{R}^{\mathrm{h}}} \sum_{s \in \mathcal{S} \setminus |\mathcal{S}|} c_k^{\mathrm{o}} \cdot o_{k,s}^{\mathrm{net}} \tag{B3}$$

Appendix C. Distribution Fitting Results

We used simul8's StatFit tool (Domonkos (2010)) and the built-in Anderson-Darling test to check whether the empirical recovery time and the capacity distribution fit to lognormal distributions. The null-hypothesis is that the empirical distribution is equal to the theoretical distribution. In both cases, the outcome was 'do not reject'. The recovery time and remaining resource capacity distributions are shown in Figure C1(a) and (b), respectively.



(a) (b)

Figure C1. Result of fitting the recovery time (a) and capacity distribution (b)

References

Ahmadi-Javid A, Jalali Z and Klassen K (2017). Outpatient appointment systems in healthcare: A review of optimization studies. *European Journal of Operational Research* **258**(1): 3–34.

Arnolds I and Gartner D (2018). Improving Hospital Layout Planning Through Clinical Pathway Mining. *Annals of Operations Research* **263**(1-2): 453–477.

Bastos LS, Marchesi JF, Hamacher S and Fleck JL (2018). A mixed integer programming approach to the patient admission scheduling problem. *European Journal of Operational Research* Online first. https://doi.org/10.1016/j.ejor.2018.09.003.

Bilgin B, Demeester P, Misir M, Vancroonenburg W and Van den Berghe G (2012). One hyper-heuristic approach to two timetabling problems in health care. *Journal of Heuristics* **18**(3): 401–434.

Burdett RL and Kozan E (2018). An integrated approach for scheduling health care activities in a hospital. *European Journal of Operational Research* **264**(2): 756–773.

Burdett RL, Kozan E, Sinnott M, Cook D and Tian YC (2017). A mixed integer linear programing approach to perform hospital capacity assessments. *Expert Systems with Applications* **77**: 170–188.

Ceschia S and Schaerf A (2016). Dynamic patient admission scheduling with operating room constraints, flexible horizons, and patient delays. *Journal of Scheduling* **19**(4): 377–389.

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

Domonkos T (2010). Computer simulation as a tool for analyzing and optimizing real-life processes. *Management Information Systems* **5**(1): 13–18.

Eijkemans M, van Houdenhoven M, Nguyen T, Boersma E, Steyerberg E and Kazemier G (2010). Predicting the unpredictable: A new prediction model for operating room times using individual characteristics and the surgeon's estimate. *Anesthesiology* **112**(1): 41–49.

Gartner D (2015). Scheduling the hospital-wide flow of elective patients. *Springer Lecture Notes in Economics and Mathematical Systems* Heidelberg.

Gartner D and Kolisch R (2014). Scheduling the hospital-wide flow of elective patients. *European Journal of Operational Research* **223**(3): 689–699.

Gartner D, Kolisch R, Neill DB and Padman R (2015). Machine Learning Approaches for Early DRG Classification and Resource Allocation. *INFORMS Journal on Computing* **27**(4): 718–734.

Gartner D and Padman R (2016). Length of Stay Outlier Detection in Pediatrics. *International Conference on Information Systems (ICIS), Dublin* .

Gartner D and Padman R (2017a). E-HOSPITAL – A Digital Workbench for Hospital Operations and Services Planning using Information Technology and Algebraic Languages. *Studies in Health Technology and Informatics* **245**: 84–88.

Gartner D and Padman R (2017b). Handbook of Research on Healthcare Administration and Management, chapter Mathematical Programming and Heuristics for Patient Scheduling in Hospitals: A Survey. IGI Global, 627–645.

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

Kazemian P, Sir M, Van Oyen M, Lovely J, Larson D and Pasupathy K (2017). Coordinating clinic and surgery appointments to meet access service levels for elective surgery. *Journal of Biomedical Informatics* **66**: 105–115.

Leeftink AG, Bikker IA, Vliegen IMH and Boucherie RJ (2018). Multi-disciplinary planning in health care: A review. *Health Systems* : 1–24Online first. Doi: 10.1080/20476965.2018.1436909.

Marynissen J and Demeulemeester E (2019). Literature review on multi-appointment scheduling problems in hospitals. *European Journal of Operational Research* **272**(2): 407–419.

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

Neumann K, Schwindt C and Zimmermann J (2003). Project scheduling with time windows and scarce resources. Springer, Berlin, 2nd edition.

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

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

Sharma MJ and Yu SJ (2009). Benchmark optimization and attribute identification for improvement of container terminals. *European Journal of Operational Research* **201**(2): 568–580.

Turhan A and Bilgen B (2017). Mixed integer programming based heuristics for the Patient Admission Scheduling problem. *Computers & Operations Research* **80**: 38–49.

Wauters T, Kinable J, Smet P, Vancroonenburg W, Van den Berghe G and Verstichel J (2016). The Multi-Mode Resource-Constrained Multi-Project Scheduling Problem. *Journal of Scheduling* **19**(3): 271–283.