

Schwarz, Karsten; Römer, Michael; Mellouli, Taïeb

## Article

# A data-driven hierarchical MILP approach for scheduling clinical pathways: a real-world case study from a German university hospital

Business Research

## Provided in Cooperation with:

VHB - Verband der Hochschullehrer für Betriebswirtschaft, German Academic Association of Business Research

*Suggested Citation:* Schwarz, Karsten; Römer, Michael; Mellouli, Taïeb (2019) : A data-driven hierarchical MILP approach for scheduling clinical pathways: a real-world case study from a German university hospital, Business Research, ISSN 2198-2627, Springer, Heidelberg, Vol. 12, Iss. 2, pp. 597-636,  
<https://doi.org/10.1007/s40685-019-00102-z>

This Version is available at:

<https://hdl.handle.net/10419/233149>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*


*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>

# A data-driven hierarchical MILP approach for scheduling clinical pathways: a real-world case study from a German university hospital

Karsten Schwarz<sup>1</sup>  · Michael Römer<sup>2</sup> ·  
Taïeb Mellouli<sup>2</sup>

Received: 4 February 2016 / Accepted: 10 September 2019 / Published online: 19 September 2019  
© The Author(s) 2019

**Abstract** Facing economic pressure and case-based compensation systems, hospitals strive for effectively planning patient hospitalization and making efficient use of their resources. To support this endeavor, this paper proposes a flexible hierarchical mixed-integer linear programming (MILP)-based approach for the day-level scheduling of clinical pathways (CP). CP form sequences of ward stays and treatments to be performed during a patient’s hospitalization under consideration of all relevant resources such as beds, operating rooms and clinical staff. Since in most hospitals CP-related information needed for planning is not readily available, we propose a data-driven approach in which the structure of the CP to be scheduled including all CP-related constraints is automatically extracted from standardized hospital billing data available in every German hospital. The approach uses a flexible multi-criteria objective function considering several patient- and hospital-related aspects which makes our approach applicable in various scenarios. Furthermore, in contrast to other approaches, it considers several practically relevant aspects ensuring the implementability of the scheduling results such as multiple ward stays per hospitalization and gender-separated room assignments. Regarding the treatment resources such as operation rooms and clinical staff, it considers the eligibility of resources for treatments based on information such as special

---

✉ Karsten Schwarz  
karsten.schwarz@medizin.uni-halle.de

Michael Römer  
michael.roemer@wiwi.uni-halle.de

Taïeb Mellouli  
mellouli@wiwi.uni-halle.de

<sup>1</sup> Dorothea Erxleben Learning Center, Martin-Luther-University Halle-Wittenberg, Halle, Germany

<sup>2</sup> Department of Management Information Systems and Operations Research, Martin-Luther-University Halle-Wittenberg, Halle, Germany

equipment or qualification and represents complex resources individually to avoid disaggregation problems. To allow solving the resulting complex and large-scale scheduling problem for realistically dimensioned problem instances, we propose a hierarchical two-stage MILP approach involving carefully designed anticipation components in the first-stage model. We evaluate our approach in a case study with real-world data from a German university hospital showing that our approach is able to solve instances with a planning horizon of 1 month exhibiting 1088 treatments and 302 ward stays of 286 patients. In addition to comparing our approach to a monolithic MILP approach, we provide a detailed discussion of the scheduling results for two practically motivated scenarios.

**Keywords** Clinical pathways · Scheduling · MILP · Health care planning · Patient flow · Data driven · Case study

## 1 Introduction

In Germany, hospitals cause around 25% of the total costs of the health care system (Destatis 2014). To reduce these costs of about 76bn Euro per year, the German Diagnosis-Related Groups (DRG) system was established in 2003. In contrast to the earlier length-of-stay-based reimbursement system, hospitals now receive a diagnosis-dependent lump sum per case. As a consequence, hospitals have a strong incentive to reduce their patients' lengths of stay and to effectively manage clinical processes and resources.

At the time of writing, however, every third hospital is still unprofitable (Augurzky 2015) which, together with demographic change, yields a great need for increasing the efficiency of hospitals. According to Villa et al. (2009), clinical pathways (CP) are among the most promising instruments for achieving an efficient utilization of both human and physical hospital resources. CP are specific sets of time-constrained treatments and ward stays to be performed between a patient's admission and discharge to cure a certain disease. Besides increasing the transparency and the standardization of medical processes, CP are instrumental in hospitalization scheduling and controlling (Jacobs 2007). Despite this often-stated potential, however, CP and patient flow scheduling are still waiting to form an integral part of day-to-day practice in many hospitals.

According to Roeder et al. (2003), Kirschner et al. (2007), Küttner and Roeder (2007), and Salfeld et al. (2009), an important obstacle for a widespread utilization of CP is the fact that manually creating CP forms a tedious endeavor. In addition, since CP are mostly used in a descriptive way, only few CP formalisms discussed in the literature enable both the automatic extraction of CP structures from hospital data and their employment for scheduling. Furthermore, many important parameters affecting pathway schedules such as emergency patient arrivals, (exact) treatment durations, resource availability or complications arising in surgeries are subject to uncertainty. Moreover, given that the pathway-scheduling problem has to consider complex constraints as well as all relevant resources, it constitutes a complex large-scale optimization problem. As a consequence, many pathway-scheduling

approaches rely on simplifications and operate on a high level of aggregation which makes putting the resulting schedules into practice difficult: The approach from Saadani et al. (2014), for example, completely ignores bed resources. Finally, while hospital scheduling typically considers multiple objectives related to economics, patients and staff members, many approaches focus on one-dimensional performance indicators, typically based on cost minimization or profit maximization (Hulshof et al. 2012).

Motivated from a case study with a German university hospital, this paper addresses most of the issues discussed above and proposes a tactical pathway-scheduling approach intended to support hospital case managers responsible for the coordination and scheduling of treatments and ward stays. The main contributions of this paper can be summarized as follows.

First, we propose a data-driven approach to pathway scheduling. Our approach is based on scheduling relevant pathway information automatically extracted from standardized hospital billing data available for every hospital in Germany using the pathway-mining approach introduced by Helbig et al. (2015). As a result, one of the most time-consuming data preparation steps necessary for pathway scheduling—defining pathways and their constraints—is automated to a large extent. To the best of our knowledge, the resulting approach is the first to combine process mining techniques with optimization-based decision support for scheduling clinical pathways.

Second, we propose a day-level pathway-scheduling approach considering all pathway constraints and hospital resources on an adequate level of detail which is flexible enough to consider multiple scheduling criteria: it considers multiple ward stays during hospitalization and gender-separated room assignments. Moreover, it explicitly accounts for resource eligibility constraints for treatments, e.g., based on specific feature and skill requirements. In addition, it allows considering important resources individually while aggregating identical resources if appropriate. Finally, to support scheduling pathways for a broad range of possible objective structures, the objective function in our approach involves a weighted combination of multiple hospital- and patient-related criteria including fairness and workload-balancing criteria.

Third, to tackle the resulting complex large-scale scheduling problem for realistic problem instances involving up to 300 patients with a planning horizon of a full month, we propose a two-stage mixed-integer linear programming (MILP) approach based on a hierarchical decomposition: The first-stage model determines the admission dates and schedules complex treatments with a high duration. Given instructions from the first model, the second model schedules the ward stays, assigns all patients to gender-separated rooms and schedules the remaining treatments. To align the first-stage instructions with the second-stage requirements, the first-stage model involves anticipation components to account for the gender separation requirement and to consider pathway constraints involving the full set of treatments.

Fourth, we evaluate the overall approach including the automatic CP extraction in a set of experiments conducted with real-world data from a German university hospital involving 286 elective patients and 1088 treatments. On the one hand, we evaluate our hierarchical approach in comparison with a monolithic mixed integer

linear programming (MILP) approach: we compare solution times as well as solution quality of both approaches; furthermore, we investigate the effect of employing different degrees of anticipation within the hierarchical approach. On the other hand, we provide a detailed evaluation of two scheduling scenarios: in the first scenario, our approach is used to establish balanced resource utilization throughout the planning horizon; in the second scenario, the main goal is to obtain a high resource utilization at the beginning of the planning horizon.

Note that in this paper, we do not explicitly address the stochastic character of the pathway-scheduling problem: on the one hand, our approach only deals with scheduling elective patients for which most planning-relevant information is known ahead. Nonetheless, one of the goals of the planning approach is to establish balanced resource utilization and resource buffers to ensure that emergency cases can be handled adequately. On the other hand, our approach employs conservative point estimates for uncertain treatment times to obtain schedules which are robust with regard to treatment time variations. Like the other authors dealing with clinical pathway planning (see e.g., Gartner and Kolisch 2014 and Burdett and Kozan 2018), we assume that additional uncertainties such as absence of medical staff, resource breakdowns or stochastic patient arrivals are addressed by embedding our pathway-scheduling approach in a rolling planning process.

The remainder of this paper is structured as follows. Section 2 provides an overview of the rich literature dealing with offline operational scheduling in hospitals and identifies possible objectives, efficient modeling techniques and aspects to be considered in clinical pathway scheduling. Section 3 introduces the constraint-based CP concept according to which the scheduling is carried out and sketches the automated CP mining approach mentioned above. In Sect. 4, we introduce our hierarchical MILP approach consisting of an aggregated first-stage model for scheduling complex treatments and determining the admission day and a detailed second-stage model for scheduling the remaining treatment and assigning patients to rooms. In Sect. 5, we explain the real-world dataset from a university hospital used in our computational experiments. In addition, we illustrate the results from mining CP for the hospital's largest department. Section 6 presents the results from using our MILP approach for scheduling the extracted CP for all elective patients of the department of urology. To demonstrate the effects of varying the weights in the multi-criteria objective function, we discuss two scenarios: in the first scenario, the resource allocation is balanced across the planning horizon; in the second scenario, resources are preferably allocated at the beginning of the horizon. Section 7 concludes the paper and points to further research opportunities.

## 2 Related work

We employ the taxonomy of planning decisions in health care introduced by Hulshof et al. (2012) to structure the following literature review. According to this taxonomy, the problem considered in this paper can be characterized as an offline operational planning problem: complete CP for elective patients are scheduled within a planning horizon of about 1 month considering activities at the level of

individual patients and resources. Since our approach takes all kinds of hospital resources into account, the following literature review does not only regard CP-related literature but also considers more general hospital resource scheduling approaches to identify aspects relevant for our research. Since the assignment of staff to shifts and inventory management issues are not in the scope of this paper, articles dealing with these topics are not included in the following review. Furthermore, the review is restricted to articles involving mathematical programming as solution technique published after 2005. Table 1 shows the related work classified by the combinations of considered resources and treatments.

Since surgeries form the greatest source of income for most hospitals (Denton et al. 2007) and at the same time operating rooms (OR) constitute the most expensive type of resource using more than 10% of a typical hospital budget (Jebali et al. 2006; Chaabane et al. 2008), there is a rich body of literature dealing with surgery scheduling. The main goals of surgery scheduling are to reduce costs (overtime costs, penalties for idle time or fixed costs for opening an OR) and to increase the utilization of the available OR. This is achieved by either allocating surgeries to a given set of OR (Lamiri et al. 2007; Lamiri et al. 2008a, b; Fei et al. 2009; Denton et al. 2010; Riise and Burke 2011; Meskens et al. 2013), sequencing a set of given surgeries (Denton et al. 2007; Cardoen et al. 2009a; Cardoen and Demeulemeester 2011) or both (Jebali et al. 2006; Testi and Tãnfani 2009; Roland et al. 2010; Batun et al. 2011; Marques et al. 2012; Clavel et al. 2017). In most cases, mathematical programming, heuristics, column generation or combinations of these methods are used to find a good or even optimal solution with respect to a typically multi-criteria objective function. With regard to the scope of the present article, the main insights from the surgery scheduling literature can be summarized as follows: Overtime should be avoided and the well-being of the medical staff should be considered (Roland et al. 2010; Meskens et al. 2013), a flexible pooling strategy of OR has significant benefits (Batun et al. 2011), the eligibility of

**Table 1** Overview of related work

Resources	Related articles
Operating Room	Jebali et al. (2006), Denton et al. (2007), Lamiri et al. (2007), Chaabane et al. (2008), Lamiri et al. (2008a, b), Cardoen et al. (2009a), Fei et al. (2009), Testi and Tãnfani (2009), Denton et al. (2010), Roland et al. (2010), Batun et al. (2011), Cardoen and Demeulemeester (2011), Marques et al. (2012) and Clavel et al. (2017)
Beds	Demeester et al. (2010), Ceschia and Schaerf (2011), Schmidt et al. (2013), Helm and Van Oyen (2014) and Vancroonenburg et al. (2014)
Treatments	Vlah Jerić and Figueira (2010, 2012) and Schimmelpfeng et al. (2012)
Operating room and beds	Pham and Klinkert (2008), Cardoen et al. (2009b), Augusto et al. (2010), Fei et al. (2010), Chow et al. (2011), Banditori et al. (2013), Sun et al. (2013), Vancroonenburg et al. (2013), Ceschia and Schaerf (2014) and Li et al. (2015)
Clinical pathways	Vissers (2005), Conforti et al. (2011), Helbig (2011), Gartner and Kolisch (2014) and Saadani et al. (2014)

operating rooms for surgeries may depend on special operating room equipment (Denton et al. 2007) and the patients' welfare should be taken into account (Testi and Tànfani 2009). Regarding the solution methods, carefully decomposing the problem into solving multiple small problems often yields a good performance in terms of solution time without having a significantly negative impact on solution quality (Jebali et al. 2006; Lamiri et al. 2007; Lamiri et al. 2008a, b; Fei et al. 2009; Denton et al. 2010; Batun et al. 2011; Marques et al. 2012). For a recent survey on OR scheduling, see Samudra et al. (2016).

As stated by Helm and Van Oyen (2014), a proper bed management is necessary to avoid having to dismiss patients because of blocked beds, surgical cancelations and operational chaos. In the literature, two kinds of bed management problems are discussed both of which typically consider medical needs as well as patient preferences: elective admission scheduling and the patient-to-room assignment. Elective admission scheduling problems are solved, e.g., by local search heuristics (Ceschia and Schaerf 2011), MILP (Helm and Van Oyen 2014) or random forest models (Schmidt et al. 2013). Demeester et al. (2010) solve the patient-to-room assignment problem using a tabu search heuristic. Vancroonenburg et al. (2014) develop an integrated model to solve both problems at once to increase both planning flexibility and operational efficiency. Gender-separated room assignment is typically considered as a crucial issue, see Demeester et al. (2010, Ceschia and Schaerf (2011), Schmidt et al. (2013), and Vancroonenburg et al. (2014).

According to Schimmelpfeng et al. (2012), scheduling treatments (encompassing not only surgeries but all types of medical activities) manually often leads to an inefficient resource allocation and can have negative effects on quality of care and patient satisfaction. To avoid this, Vlah Jerić and Figueira (2010) develop a decision support system (DSS) to support the construction of a daily schedule of medical treatments considering available resources and other criteria. They use a scatter search heuristic to construct a set of Pareto-optimal solutions among which the user can interactively choose. Another approach by the same authors presented in Vlah Jerić and Figueira (2012) determines daily treatment schedules considering the availability of medical equipment and physicians. Their approach is based on a multi-objective binary programming formulation for which different solution approaches such as variable neighborhood search, scatter search and non-dominated sorting genetic algorithms are compared. Schimmelpfeng et al. (2012) outline a conceptual framework for a DSS to support the scheduling process of treatments in a rehabilitation hospital. To handle realistic problem dimensions, a hierarchical approach is developed using a daily model for aggregated long-term scheduling and an intra-day model (either time- or resource-oriented) for short-term scheduling.

As argued by Chow et al. (2011), trying to achieve a high OR utilization without considering further surgery-related resources often results in issues such as staff overtime, surgical cancelations and long surgical waiting times. In particular, scheduling surgeries without taking the availability of recovery beds into account often leads to problems. To avoid this issue, several authors propose to simultaneously schedule both the surgery and the following stay in a recovery bed or on a ward. For instance, Pham and Klinkert (2008) propose to allocate hospital resources to individual surgical cases divided into preoperative,

perioperative (= surgery), postoperative intensive care unit (ICU) modes and formulate this problem as a MILP multi-mode blocking job shop model. Cardoen et al. (2009b) solve a surgical case scheduling problem in a day-care facility. They formulate a multi-objective model to minimize peaks in recovery bed usage, the occurrence of recovery overtime and violations of staff and patients' preferences. Their approach is based on a combination of column generation and dynamic programming. To analyze the impact of recovery in OR when no recovery bed is available, Augusto et al. (2010) use a Lagrangian relaxation-based approach. Their results show a high benefit of this improved flexibility in resource usage, in particular in case of high demands for recovery beds. Fei et al. (2010) construct weekly surgery schedules for OR considering recovery beds to maximize resource utilization while minimizing overtime and idle time between surgeries. They use a hierarchical two-stage solution approach which first computes the date of the surgeries based on a set partitioning formulation solved by a column generation technique. In the second step, taking into account the availability of recovery beds the daily surgery scheduling problem in which the sequence of surgeries is determined is formulated as a flow-shop problem and solved by a hybrid genetic heuristic. The authors show that their overall approach allows reducing idle- and overtime increasing resource utilization. Chow et al. (2011) use Monte Carlo Simulation to predict bed requirements from historical data. Based on this, they formulate a MILP to reduce peaks in bed occupancy by scheduling surgeon blocks and patient types. Another simulation optimization approach for the surgery scheduling problem aiming at a maximal patient throughput is developed by Banditori et al. (2013). The authors first use a MILP model to determine the number of cases to be treated by surgery groups in each time slot of a month. The model is then fine-tuned with a simulation approach to obtain both robust and easy-to-implement schedules. Sun et al. (2013) propose a weekly scheduling approach to achieve a high OR utilization. The authors formulate a MILP taking into account limited key resources like ICU-beds and workload of surgeons. Vancroonenburg et al. (2013) incorporate OR and gender separation constraints in their approach for the elective patient admission scheduling problem. Their goal is to determine how scheduling surgical and non-surgical admissions impacts room assignments at wards. Ceschia and Schaerf (2014) address a similar patient admission problem considering OR utilization constraints, gender-separated room assignment, a flexible planning horizon and the notion of patient delay using a local search heuristic. Li et al. (2015) develop a MILP-based lexicographic goal programming approach of scheduling OR and beds considering competing resources such as surgical and nursing staff, anesthesiologists and recovery beds.

The most promising way to avoid bottlenecks, idle time and overtime during hospital patient flow is to schedule not only surgeries, possibly augmented with induced bed resources, but each patient's full CP. Vissers (2005) proposes a master surgery scheduling approach for CP involving cardiothoracic surgeries based on computing the patient mix with a MILP model. He divides the care process into four successive steps (medium care unit (MCU), surgery, ICU, MCU) and computes the master schedule for a planning cycle of 4 weeks taking OR, MCU-beds, ICU-beds and nursing staff for ICU-beds into account. An approach to schedule full CP of



elective patients for a week hospital<sup>1</sup> is introduced by Conforti et al. (2011). The authors show their results from scheduling clinical services (diagnostics and surgeries) as well as patient admissions to maximize the patient flow using a MILP model in a case study with 20 patients. Helbig (2011) presents a MILP model for the operational hour-based scheduling of interdisciplinary CP containing treatments, surgeries, order constraints and ward stays. The aim of the model is to minimize waiting time and to finish each CP as early as possible within the planning horizon under consideration of practical constraints like gender separation in rooms and a fair distribution of waiting time among all patients. The paper reports results from experiments with artificial small-scale problem instances with up to four patients. A patient flow planning approach to maximize the length of stay (LOS)-based contribution margin for DRG-based reimbursement policies is developed by Gartner and Kolisch (2014). The authors formulate two MILP models to schedule CP, one with fixed and one with variable patient admission dates. They experiment with these models in a static and in a dynamic setting in which the MILP is embedded in a rolling horizon approach. The results dealing with a planning horizon of 28 days and 150 patients show the potential to achieve a higher contribution margin and a significantly reduced time between admission and surgery compared to given manual solutions. Saadani et al. (2014) propose a MILP-based CP scheduling approach in which treatments using different resources are scheduled to minimize the patients' length of stay. Their CP contain multiple treatments but no bed requirements. The authors present results for an instance of 20 patients and a planning horizon of 14 days. Recently, Burdett and Kozan (2018) proposed an approach using constructive heuristics and metaheuristics for scheduling CP operating on a very high level of detail: They schedule not only the day, but also the time of day of the treatments. Their approach takes into account all major treatment resources under consideration of resource eligibility constraints; in addition, it considers multiple ward stays and bed resources on wards.

Summarizing the results of the literature review, the most promising way to improve the patient flow in hospitals is to consider the full CP of each patient. As shown above, scheduling CP involves various complex aspects to be considered by a CP scheduling approach to achieve practically useful and implementable schedules:

- Given the multitude of relevant objectives, the approach should allow to include multiple objective criteria related to both patients and resources:
  - Important patient-related objectives encompass the consideration of LOS (the main driver of schedule-related costs in a DRG system), admission day preferences and hospitalization without delays.
  - Among the resource-related objectives, minimization of overtime and idle time, establishing a balanced workload and achieving a high level of resource utilization are the most important.

---

<sup>1</sup> A week hospital is a hospital where the duration of all hospitalizations is at most 1 week.

- To obtain implementable pathway schedules, patients should be assigned to concrete rooms taking gender separation into account.
- For an adequate representation of resource requirements, scheduling should consider the eligibility of resources for treatments which depends on resource properties such as special qualifications or equipment. Furthermore, to avoid disaggregation problems as sketched in the introduction, important treatment resources such as OR should be considered individually instead of in an aggregated way.
- It is unlikely to find a monolithic approach for exactly solving a CP scheduling problem considering all of the aspects described above. However, decomposing the problem into multiple stages, if carried out carefully, can be a promising approach to achieve good results within acceptable solution times also for large-scale instances involving more than 200 patients, a 30-day planning horizon and 1000 treatments.

We are not aware of any mathematical programming-based CP scheduling approach taking into account all the aspects mentioned above. As a consequence, this paper proposes a novel hierarchical MILP approach to schedule whole CP with hospital-wide ward stays able to handle the mentioned aspects for real-world problem instances.

An additional important feature of our approach is its data-driven nature: by employing the pathway-mining approach and the constraint-based representation of CP introduced by Helbig et al. (2015), the structure of the CP to be scheduled can be automatically extracted from standardized data available in every German hospital. It becomes clear from the recent survey articles Yang and Su (2014) and Rojas et al. (2016) that process mining for clinical pathways and in healthcare in general forms a very active field of research. As described in Rojas et al. (2016), however, the results of the process mining are typically used for analyzing and improving healthcare processes using techniques from business process management. Our approach in contrast is, to the best of our knowledge, the first one tightly coupling process mining with mathematical optimization for scheduling clinical pathways.

### 3 Constraint-based CP representation

Even though the general concept of CP is well-known and widely accepted, there is no standard formalism for representing CP (Vanhaecht et al. 2006). While the formal representation of a CP is of secondary importance if it serves merely descriptive purposes, it is crucial when CP form the basis of scheduling patient flow. Conforti et al. (2011), for example, consider a CP as a set of treatments to be scheduled without imposing inter-treatment dependencies. Gartner and Kolisch (2014) represent the possible schedules for each CP on a directed graph in which the arcs form minimum time lags between the clinical activities. The CP scheduling approach proposed in this work relies on the constraint-based CP representation introduced in (Helbig et al. 2015). We give a brief overview and example of this

representation in this section; for details regarding the representation as well as regarding the pathway-mining approach, we refer the reader to (Helbig et al. 2015).

As illustrated in Fig. 1, the time scale of the pathway representation is days and a CP is considered on two levels: the level of ward stays and the level of treatments. On the level of ward stays, a CP forms a chronological sequence of ward stays  $\mathcal{S}$ . Note that while in many cases, the set of ward stays  $\mathcal{S}$  is a singleton, there are cases in which ward changes are part of the planned pathway of elective patients: as an example, in the department of urology considered in our case study, there are several cases with complex surgeries involving both a stay in an intensive care unit (ICU) and on a regular ward; in total, 6% of the cases in the dataset involve multiple ward stays.

In the CP representation, a ward stay  $s \in \mathcal{S}$  is associated with a set of eligible wards  $\mathcal{W}_s$ ; this means that for stay  $s$ , the patient may be assigned to one of the wards in  $w_s$ . In Fig. 1, for example, the ward stay  $s_2$  may either be assigned to ward  $w_2$  or to ward  $w_3$ . A ward stay is characterized by an admission event  $v_s^{ad}$  and a discharge event  $v_s^{dis}$ . The admission event  $v_{s_1}^{ad}$  of the first stay  $s_1$  corresponds to the hospital admission and always takes place at the first day of the CP. All other ward stay events have feasible time intervals indicating when they can be scheduled (e.g., the discharge  $v_{s_1}^{dis}$  can be scheduled from day 3 to day 6, see Fig. 1). The number of days between admission and discharge on a ward forms the length of stay (LOS) on that ward. For every ward stay, the LOS is constrained by a minimum and a maximum number of days. During hospitalization, each ward stay requires a bed resource and it is assumed that there are no gap days between two adjacent ward stays.

On the second level of the CP representation, a treatment  $t$  corresponds to a surgery or any other clinical service to be performed to cure a disease. A CP

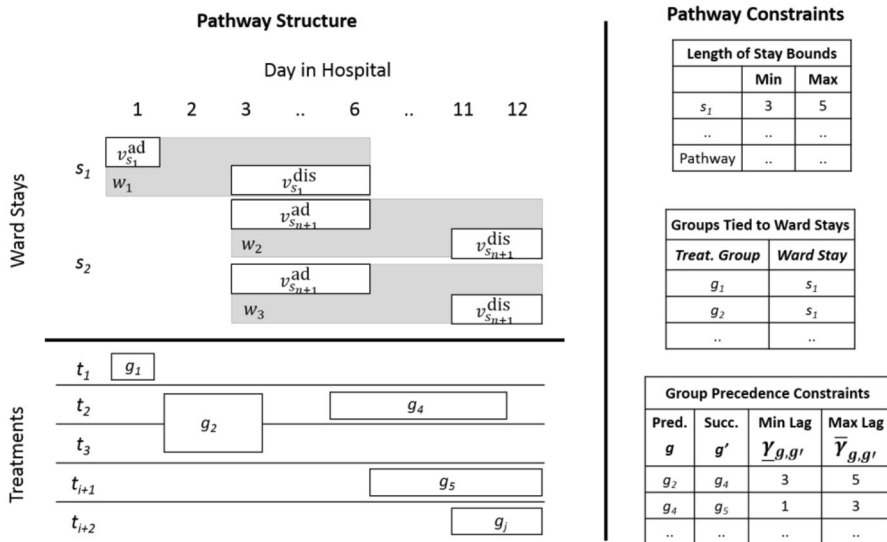


Fig. 1 Constraint-based CP concept

contains a set  $\mathcal{T}$  of different treatments. Each of these treatments lasts at most 1 day and can only be performed if the required amount of all necessary resource types (e.g., surgeons, special rooms, theater nurses or the patient) is available. To facilitate scheduling, treatments are clustered into treatment groups; the treatments within the same group  $g$  have to be scheduled at the same day (see e.g.,  $t_2$  and  $t_3$  in Fig. 1). Some treatment types may occur in multiple groups which means that they have to be performed once per group ( $t_2$ , for example, occurs in two groups). To ensure medically correct schedules, each treatment group  $g$  has a feasible time interval  $[\underline{\delta}_g; \bar{\delta}_g]$  relative to the first admission event. In Fig. 1, this interval corresponds to the length of the bar associated with the treatment group (e.g.,  $g_2$  has the feasible interval [2;3]). In addition, a medically correct order can be enforced by precedence constraints imposing a minimum and a maximum time lag  $[\underline{\gamma}_{g,g'}; \bar{\gamma}_{g,g'}]$  between two related treatment groups  $g$  and  $g'$  (e.g., in Fig. 1,  $g_5$  needs to fall into the interval [1;3] relative to the scheduled day of  $g_4$ ). Moreover, a group can be associated with a certain ward stay  $s$  which means that the group can only be scheduled during that ward stay (e.g.,  $g_1$  and  $g_2$  from Fig. 1 can only be scheduled during the first ward stay). Note that by grouping treatments as described, many CP-related constraints are aggregated and in some cases even considered implicitly. On the one hand, this contributes to a tractable size of the MILP models presented in the next section, on the other hand, the modeling itself is facilitated by the grouping.

#### 4 A hierarchical MILP approach for scheduling clinical pathways

The pathway-scheduling problem considered in this paper consists of determining the patient's admission and discharge days, the start and end days of the ward stays and the days on which treatments are performed. The scheduling accounts for precedence relations between treatments and considers all critical hospital resources such as beds, medical staff as well as treatment and operating rooms. Bed resources are considered on the level of rooms, accounting for the requirement of gender separation. Treatments are associated with resource requirements. It is assumed that for each requirement there is a set of eligible resources and a given amount of resource time needed for the treatment. Each resource, in turn, is associated with a time capacity for each given day. Identical resources may be considered in an aggregated way; however, to avoid disaggregation problems, important complex resources such as operating rooms are considered individually.

With respect to the level of detail considered, the pathway-scheduling problem addressed in this paper is situated between the problems considered in Gartner and Kolisch (2014) and in Burdett and Kozan (2018): both the problems considered in Gartner and Kolisch (2014) and in this paper schedule treatments on the time-aggregated level of days, but in contrast to the problem addressed here, Gartner and Kolisch (2014) do not consider multiple ward stays, bed assignment to gender-separated rooms and resource eligibility constraints. Burdett and Kozan (2018), in contrast, consider multiple ward stays and resource eligibility constraints such that the level of detail for representing resources is similar to our work. However, the

problem addressed in Burdett and Kozan (2018) deals with the exact timing of treatments instead of only determining the day on which a treatment is performed.

Gartner and Kolisch (2014) point out that the pathway-scheduling problem they address is related to the resource-constrained project scheduling problem (RCPSP), with the important difference that treatments are assumed to take a fraction of a the time unit (in this case: a day) while in project scheduling, activities are typically assumed to take multiple units of time. Using the terminology from Artigues et al. (2015), the model presented by Gartner and Kolisch (2014) can be characterized as a time-indexed formulation, that is, the main decision variables for scheduling activities are binary variables with a day index. The authors assume that each resource requirement is associated with a single set of aggregated resources. As a result, no treatment-to-resource assignment is needed and the resource capacity constraint can be formulated using the time-indexed treatment scheduling variables. The resulting model instances are fairly compact and can be solved to optimality by standard MILP solvers even for realistically sized instances.

Using the analogy to RCPSP, extending the problem addressed in Gartner and Kolisch (2014) by the consideration of treatment-specific resource eligibility is very similar to the extension of the RCPSP to a project staffing and scheduling problem in which resources are considered as flexible or multi-skilled and each task is associated with a set of eligible resources. As discussed in Correia and Saldanha-da-Gama (2015), this setting adds a whole new dimension to the RCPSP since tasks are not only scheduled but also explicitly assigned to resources. Incorporating resource eligibility in MILP formulations for the RCPSP typically results in additional variables and constraints, see, e.g., the formulation proposed in Correia and Saldanha-da-Gama (2015) and, as also discussed therein, makes the problems much harder to solve. Note that the modeling framework for project staffing and scheduling problems proposed in Correia and Saldanha-da-Gama (2015) is based on natural-date variables, that is, on variables representing start times of tasks. Our formulation presented below employs both natural-date variables (e.g., for the start day of ward stays) and time-indexed variables (e.g., for resource assignments). Note that Burdett and Kozan (2018) do not provide a MILP formulation of their very detailed CP scheduling problem—their solution approach uses constructive heuristics and metaheuristics.

Instead of resorting to metaheuristic approaches to deal with the computational complexity of the problem considered in this paper, similar to Schimmelpfeng et al. (2012), we employ a two-stage MILP approach which can be viewed as a constructional hierarchical decomposition in the terminology of hierarchical planning, see, e.g., (Schneeweiss 1998). The first stage determines the “cornerstones” of the clinical pathway, that is, the patient admission and discharge dates as well as the dates for the complex key treatments of each clinical pathway. Considering the first-stage decisions as fixed instructions, the second stage solves the full problem sketched above.

To obtain a high-quality solution for the full problem, the first-stage model anticipates major aspects of the full planning problem using anticipation model components obtained by the means of aggregation and relaxation. While the first stage does not determine the ward stays, it anticipates ward stay scheduling under

consideration of the gender separation requirement on the level of aggregated room types. Furthermore, while it only schedules complex treatments (under consideration of all required treatment resources), to make this scheduling consistent with the pathway constraints involving all treatments, the first-stage model also involves all remaining treatment groups (neglecting the respective resource requirements). Before presenting the MILP formulations for the first-stage and the second-stage model, we give a short overview of the notational conventions used in the presentation. A full list of all symbols used in the models is given in the Appendix; in addition, all symbols are explained in the description of the models.

#### 4.1 Notation conventions and variable domains

We use calligraphic symbols for representing sets, in some cases refined with descriptive superscripts. As an example,  $\mathcal{P}$  denotes the set of patients and  $\mathcal{P}^m$  denotes the set of male patients. Descriptive superscripts are also used for variables and parameters; subscripts are reserved for indices. Indices and parameters are in lowercase. Greek letters are used for path-relative time information (see Sect. 3).

Decision variables are capitalized and we utilize the following conventions to enhance the readability of the models: the letter  $X$  is used for variables for which the value corresponds to a day in the planning period, that is, the domain of these variables is the set  $\mathcal{D} = \{1, \dots, |\mathcal{D}|\}$ . Superscripts are used to refer to specific groups of variables, e.g.,  $X_{p,s}^{ws}$  denotes the start day of ward stay  $s$  of patient  $p$ . Day-indexed variables are mostly denoted with  $Y$ . For example, the binary variable  $Y_{p,w,d}^{\text{ward}}$  indicates whether patient  $p$  stays at ward  $w$  at day  $d$  or not. Bookkeeping variables representing a duration, a slack or a violation of a soft constraint are denoted with  $U$ ; again specified with a descriptive superscript and possibly with an additional superscript signifying the direction of a violation. As an example, the variable  $U_p^{\text{ad}+}$  represents a positive violation from the desired admission date of patient  $p$ . If not specified otherwise, the bookkeeping variables are continuous non-negative variables. The letter  $Z$  is used for integer variables mostly representing assignment decisions—for example, the indicator variable  $Z_{p,s,q}^{\text{room}}$  takes the value of 1 if patient  $p$  stays in room  $q$  throughout her ward stay  $s$ .

#### 4.2 First-stage model

In the following exposition, we present the first-stage model used to determine the admission and discharge dates of the patients as well as the dates of the (complex) key treatments in the clinical pathways. Note that all other decisions considered in the following model such as the start and end days of the ward stays, the aggregated gender-separated room assignments and the scheduled dates of the non-complex treatment groups are not transferred to the second-stage model as instructions but merely serve the purpose of anticipation.

### 4.2.1 Scheduling admission, ward stays and discharge

For each ward stay  $s$  from the ordered set  $\mathcal{S}_p = \{1 \dots |\mathcal{S}_p|\}$  of ward stays of patient  $p$ , the decision variable  $X_{p,s}^{ws}$  represents the start day of  $s$ . The set of constraints (1) both compute the length  $U_{p,s}^{los}$  of each ward stay and establish the correct sequencing of each patient’s ward stays. Note that in the constraint sets associated with the last ward stay  $|\mathcal{S}_p|$  of patient  $p$ , the index  $|\mathcal{S}_p| + 1$  arises as a subscript of the ward stay start variable. This variable can be interpreted as the first day after the end of hospitalization. The set of constraints (2) enforces the hard lower bound  $\underline{b}_{p,s}^{los}$  as well as the soft upper bound  $\bar{b}_{p,s}^{los}$  along with the upper bound violation variable  $U_{p,s}^{los+}$  for the length of stay of ward stay  $s$  of patient  $p$ :

$$X_{p,s+1}^{ws} - X_{p,s}^{ws} = U_{p,s}^{los} \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p \tag{1}$$

$$\underline{b}_{p,s}^{los} \leq U_{p,s}^{los} \leq \bar{b}_{p,s}^{los} + U_{p,s}^{los+} \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p \tag{2}$$

Every patient  $p$  has a desired admission day  $d_p^{desAd}$ . In the set of constraints (3), the deviation of the scheduled admission day  $X_{p,1}^{ws}$  from this day is registered by the variables  $U_p^{ad-}$  and  $U_p^{ad+}$  both of which are bound by the maximum allowed deviation  $\bar{b}^{devAd}$  enforced by the constraint set (4):

$$X_{p,1}^{ws} = d_p^{desAd} - U_p^{ad-} + U_p^{ad+} \quad \forall p \in \mathcal{P} \tag{3}$$

$$0 \leq U_p^{ad-} \leq \bar{b}^{devAd}, 0 \leq U_p^{ad+} \leq \bar{b}^{devAd} \quad \forall p \in \mathcal{P} \tag{4}$$

### 4.2.2 Aggregated consideration of bed resources and gender separation

For anticipation purposes, the first-level model considers bed resources as well as the gender separation requirement on the level of aggregated room types. To model these requirements, the day-indexed variables  $Y_{p,s,w,d}^{ward}$  indicate whether on day  $d$ , patient  $p$  occupies a bed on ward  $w$  during ward stay  $s$  on one of the eligible wards in set  $\mathcal{W}_{p,s}$ . The three sets of constraints (5–7) link the bed occupation variables  $Y_{p,s,w,d}^{ward}$  to the scheduling of the ward stays: Constraint (5) ensures that the number of bed occupation days equals the length of the ward stays; constraints (6) and (7) force the bed occupation days to fall into the scheduled period of the corresponding ward stay:

$$\sum_{w \in \mathcal{W}_{p,s}} \sum_{d \in D} Y_{p,s,w,d}^{ward} = U_{p,s}^{los} \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p, \tag{5}$$

$$X_{p,s}^{ws} \leq d + M \left( 1 - \sum_{w \in \mathcal{W}_{s,p}} Y_{p,s,w,d}^{ward} \right) \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p, d \in D, \tag{6}$$

$$X_{p,s+1}^{ws} - 1 \geq d - M \left( 1 - \sum_{w \in \mathcal{W}_{s,p}} Y_{p,s,w,d}^{\text{ward}} \right) \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p, d \in \mathcal{D}. \tag{7}$$

To ensure that a patient is assigned to a single ward during the full ward stay, we introduce the binary ward assignment variable  $Z_{p,s,w}^{\text{ward}}$  indicating whether the ward stay  $s$  of patient  $p$  is assigned to ward  $w$ . The set of bundle constraints (8) ensures that this assignment is unique; constraints (9) ensure that the bed occupation variables only take a value of one for the assigned ward:

$$\sum_{w \in \mathcal{W}_{p,s}} Z_{p,s,w}^{\text{ward}} = 1 \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p, \tag{8}$$

$$Y_{p,s,w,d}^{\text{ward}} \leq Z_{p,s,w}^{\text{ward}} \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p, w \in \mathcal{W}_{p,s}, d \in \mathcal{D}. \tag{9}$$

While the first-stage model does not involve an assignment of patients to a concrete room, it enforces that for each day, there exists an assignment of patients to rooms respecting the gender separation requirement. To achieve this, the set of rooms  $\mathcal{Q}_w$  on ward  $w$  is partitioned into subsets  $\mathcal{Q}_{w,b}$  according to the number of beds  $b$  in the rooms; the set  $\mathcal{B}$  contains the possible number of beds in rooms. For each day  $d \in \mathcal{D}$ , the number of rooms on ward  $w$  with  $b$  beds used as female (male) room is decided by the integer decision variable  $Z_{w,b,d}^{\text{Room}}$  ( $Z_{w,b,d}^{\text{mRoom}}$ ). The constraints (10) ensure that on each day  $d$  on each ward  $w$  and for room capacity  $b$ , the total number of male and female rooms does not exceed the number  $|\mathcal{Q}_{w,b}|$  of rooms with the corresponding capacity. The constraints (11) and (12) ensure that for each ward  $w$  on each day  $d$ , the number of female (male) patients having a bed on a certain ward does not exceed the total number of beds available in the rooms assigned to female (male) patient plus the number of additional beds recorded in the variables  $U_{w,d}^{\text{fBed}+}$  and  $U_{w,d}^{\text{mBed}+}$ .

$$Z_{w,b,d}^{\text{fRoom}} + Z_{w,b,d}^{\text{mRoom}} \leq |\mathcal{Q}_{w,b}| \quad \forall w \in \mathcal{W}, b \in \mathcal{B}, d \in \mathcal{D}, \tag{10}$$

$$\sum_{p \in \mathcal{P}^f} \sum_{s \in \mathcal{S}_{p,w}} Y_{p,s,w,d}^{\text{ward}} \leq \sum_{b \in \mathcal{B}} b Z_{w,b,d}^{\text{fRoom}} + U_{w,d}^{\text{fBed}+} \quad \forall w \in \mathcal{W}, d \in \mathcal{D}, \tag{11}$$

$$\sum_{p \in \mathcal{P}^m} \sum_{s \in \mathcal{S}_{p,w}} Y_{p,s,w,d}^{\text{ward}} \leq \sum_{b \in \mathcal{B}} b Z_{w,b,d}^{\text{mRoom}} + U_{w,d}^{\text{mBed}+} \quad \forall w \in \mathcal{W}, d \in \mathcal{D}. \tag{12}$$

#### 4.2.3 Scheduling treatments considering pathway constraints

As noted in Sect. 3, our approach aggregates treatments occurring on the same day form a treatment group. In the mathematical model, the integer decision variable  $X_{p,g}^{\text{treat}}$  represents the day on which the treatment group  $g$  from the set  $\mathcal{G}_p$  of all treatment groups of a patient  $p$  is scheduled. To anticipate the pathway constraints in the first-stage model, the set  $\mathcal{G}_p$  encompasses all treatment groups (not only those containing key treatments to be scheduled in the first stage).

The scheduling of treatments is linked to the scheduling of ward stays in two ways: first, as modelled in constraints (13), for each treatment group  $g$ , there is a



time window  $[\underline{\delta}_{p,g}; \bar{\delta}_{p,g}]$  relative to the admission date of patient  $p$  associated with the variable  $X_{p,1}^{ws}$ ; violations of the time window are registered by the variables and  $U_{p,g}^{tw+}$ . Second, as modelled by constraints (14), a subset  $\mathcal{G}_{p,s}$  of the treatment groups  $\mathcal{G}_p$  of a patient  $p$  may require to be performed during the ward stay  $s$  of  $p$ :

$$X_{p,1}^{ws} + \underline{\delta}_{p,g} \leq X_{p,g}^{\text{treat}} \leq X_{p,1}^{ws} + \bar{\delta}_{p,g} + U_{p,g}^{tw+} \quad \forall p \in \mathcal{P}, g \in \mathcal{G}_p, \quad (13)$$

$$X_{p,s}^{ws} \leq X_{p,g}^{\text{treat}} \leq X_{p,s+1}^{ws} - 1 \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p, g \in \mathcal{G}_{p,s}. \quad (14)$$

In addition to the relations enforced by (13) and (14), the constraint-based pathway model described in Sect. 4 considers sequence constraints providing an interval  $[\underline{\gamma}_{g,g'}; \bar{\gamma}_{g,g'}]$  between a treatment group  $g$  and a successive treatment group  $g'$ . The set of successors of a treatment group  $g$  in the pathway of patient  $p$  subject to this type of constraint is denoted with  $\mathcal{G}_{p,g}^{\text{succ}}$ ; the respective constraints in the model can be formulated as follows:

$$X_{p,g}^{\text{treat}} + \underline{\gamma}_{p,g,g'} \leq X_{p,g'}^{\text{treat}} \leq X_{p,g}^{\text{treat}} + \bar{\gamma}_{p,g,g'} \quad \forall p \in \mathcal{P}, g \in \mathcal{G}_p, g' \in \mathcal{G}_{p,g}^{\text{succ}} \quad (15)$$

#### 4.2.4 Modelling treatment resource requirements

To perform the treatments in the aggregated treatment groups, certain resources such as operating rooms, physicians and nurses are needed. As mentioned above, the first-stage model only considers the resource requirement for a subset of treatment groups. In the following exposition, this subset is denoted with  $\mathcal{G}_p^{\text{res}}$ .

When it comes to the representation of these resource requirements, most publications dealing with pathway scheduling rely on a simple representation of resources and requirements: resources have a unique resource type (or a unique identifier) and the resource requirements are formulated in terms of time needed per resource type (or per specific resource). In reality, however, the eligibility of resources for treatment requirements depends on a set of resource properties such as qualifications (e.g., the capability of conducting a certain surgery) or features (e.g., special equipment only available in certain operating rooms). In this work, this more complex type of representing resource eligibility is modelled as follows:

We represent the resource requirements on the level of treatment groups. For each group, the requirements of all treatments in the group are aggregated. The set of resource requirements for treatment group  $g$  of patient  $p$  is denoted with  $\mathcal{K}_{p,g}$ ; the amount of resource time needed for the requirement  $k \in \mathcal{K}_{p,g}$  is given by the parameter  $a_{p,g,k}$ . A resource requirement  $k$  is further associated with a set  $\mathcal{R}_k$  of eligible resources.

For each day  $d$  and for each eligible resource  $k \in \mathcal{K}_{p,g}$  for group  $g$  of patient  $p$ , there is a binary assignment variable  $Z_{p,g,k,r,d}^{\text{res}}$  indicating whether resource  $r$  is used to cover the resource requirement  $k$  of treatment group  $g$  of patient  $p$  scheduled at day  $d$ . The set of bundle constraints (16) ensures that exactly one resource is chosen for each requirement; constraints (17) link the resource assignment variables to the treatment group scheduling variables. For each resource  $r$ , the nominal capacity at day  $d$  is given by  $\bar{b}_{r,d}^{\text{res}}$ ; using the slack variables  $U_{r,d}^{\text{ot}}$  and  $U_{r,d}^{\text{it}}$  registering the

overtime and the idle time, the corresponding resource capacity constraints are given in (18). Constraints (19) ensure that for both the amount of overtime and of idle time per resource and day is restricted by an upper bound ( $\bar{b}_r^{ot}$  and  $\bar{b}_r^{it}$ ):

$$\sum_{d \in \mathcal{D}} \sum_{r \in \mathcal{R}_k} Z_{p,g,k,r,d}^{res} = 1 \forall p \in \mathcal{P}, g \in \mathcal{G}_p^{res}, k \in \mathcal{K}_{p,g}, \tag{16}$$

$$X_{p,g}^{treat} = \sum_{d \in \mathcal{D}} \sum_{r \in \mathcal{R}_k} d Z_{p,g,k,r,d}^{res} \quad \forall p \in \mathcal{P}, g \in \mathcal{G}_p^{res}, k \in \mathcal{K}_{p,g}, \tag{17}$$

$$\sum_{p \in \mathcal{P}} \sum_{g \in \mathcal{G}_p^{res}} \sum_{k \in \mathcal{K}_{p,g}} a_{p,g,k}^{res} Z_{p,g,k,r,d}^{res} = \bar{b}_{r,d}^{res} + U_{r,d}^{ot} - U_{r,d}^{it} \quad \forall r \in \mathcal{R}, d \in \mathcal{D} \tag{18}$$

$$0 \leq U_{r,d}^{ot} \leq \bar{b}_r^{ot}; 0 \leq U_{r,d}^{it} \leq \bar{b}_r^{it} \quad \forall r \in \mathcal{R}, d \in \mathcal{D} \tag{19}$$

Note that our formulation of resource requirements and capacities given by the constraints (16–19) is valid both for individual and aggregated resources. Resources can be aggregated if they have the same properties, e.g., the same qualifications or features. For example, in our case study, the OR-nurses can be considered as an aggregated resource. In case of an aggregated resource  $r \in \mathcal{R}$ , the resource capacity  $\bar{b}_{r,d}^{res}$  per day corresponds to the sum of the capacity of all individual resources on that day.

Note, however, that in addition to adequately modelling resources with different features and skills, there is another reason for individually considering (complex) resources: the requirements for these resources often comprise multiple hours and it is not possible to switch the assigned resource during a treatment. Aggregating these resources thus easily leads to a situation where there is no feasible disaggregated allocation to individual resources for an aggregated solution.

#### 4.2.5 Objective function

The objective function of the first-stage model involves multiple patient- and resource-related objectives combined in a single function. Depending on the preferences of the decision maker, the objective function coefficients can be chosen accordingly. An important part of the objective function deals with patient delays. The constraints sets (20) and (21) compute the total delay  $U_p^{del}$  of a patient: constraint set (20) ensured that the total delay of a patient is at least as large as the number of days exceeding the medically required minimum length of stay  $b_p^{los}$ . In addition to that, the constraints (21) force the total delay to be greater than the sum of the ward stay delays  $U_{p,s}^{los+}$  and the treatment delays  $U_{p,t}^{tw+}$ . To establish fairness among the patients and to avoid huge delays of single patients, a min–max objective involving the maximum total delay  $U^{maxDel}$  of all patients is used;  $U^{maxDel}$  is forced to the maximum of all patient delays by constraints (22). Similarly, to be able to enforce a smooth resource allocation along the full planning horizon, for each resource  $r \in \mathcal{R}$ , the constraints (23) compute the maximum overtime  $U_r^{maxOt}$  and the maximum idle time  $U_r^{maxIt}$  over all days of the planning horizon:

$$U_p^{\text{del}} \geq U_p^{\text{los}} - b_p^{\text{los}} \quad \forall p \in \mathcal{P} \tag{20}$$

$$U_p^{\text{del}} \geq \sum_{s \in \mathcal{S}_p} U_{p,s}^{\text{los}+} + \sum_{t \in \mathcal{T}_p} U_{p,t}^{\text{tw}+} \quad \forall p \in \mathcal{P} \tag{21}$$

$$U^{\text{maxDel}} \geq U_p^{\text{del}} \quad \forall p \in \mathcal{P} \tag{22}$$

$$U_r^{\text{maxOt}} \geq U_{r,d}^{\text{ot}}; U_r^{\text{maxIt}} \geq U_{r,d}^{\text{it}} \quad \forall r \in \mathcal{R}, d \in \mathcal{D} \tag{23}$$

The first patient-related term in the first-stage minimization objective function (24) is the maximum total delay  $U^{\text{maxDel}}$  weighted with the penalty  $c^{\text{maxDel}}$ . The other patient-related terms involve the individual delays  $U_p^{\text{del}}$  associated with the penalty  $c^{\text{del}}$  as well as the deviations from the desired admission days  $U_p^{\text{ad}-}$  and  $U_p^{\text{ad}+}$  of each patient  $p$  weighted with the coefficient  $c^{\text{ad}}$ . The remaining elements of the objective function deal with resource-related aspects: a penalty  $c^{\text{bed}+}$  is imposed for each female and male extra bed needed per day. Regarding the treatment resources, overtime and idle time of each resource  $r \in \mathcal{R}$  are considered in the objective function two ways: first, for each resource, the maximum overtime  $U_r^{\text{maxOt}}$  and idle time  $U_r^{\text{maxIt}}$  over the full planning period are considered using the penalties  $c^{\text{maxOt}}$  and  $c^{\text{maxIt}}$ . Second, for each day  $d \in \mathcal{D}$ , the overtime  $U_{r,d}^{\text{ot}}$  and idle time  $U_{r,d}^{\text{it}}$ , weighted with the penalties  $c_d^{\text{ot}}$  and  $c_d^{\text{it}}$ , contribute to the objective function:

$$\begin{aligned} \text{min} c^{\text{maxDel}} U^{\text{maxDel}} &+ \sum_{p \in \mathcal{P}} \left( c^{\text{del}} U_p^{\text{del}} + c^{\text{ad}} \left( U_p^{\text{ad}-} + U_p^{\text{ad}+} \right) \right) \\ &+ \sum_{w \in \mathcal{W}} \sum_{d \in \mathcal{D}} c^{\text{bed}+} \left( U_{w,d}^{\text{fBed}+} + U_{w,d}^{\text{mBed}+} \right) \\ &+ \sum_{r \in \mathcal{R}} \left( c^{\text{maxOt}} U_r^{\text{maxOt}} + c^{\text{maxIt}} U_r^{\text{maxIt}} + \sum_{d \in \mathcal{D}} \left( c_d^{\text{ot}} U_{r,d}^{\text{ot}} + c_d^{\text{it}} U_{r,d}^{\text{it}} \right) \right) \end{aligned} \tag{24}$$

### 4.3 Second-stage model

Considering the instructions from the first-stage model, that is, the patient admission and discharge dates as well as the dates of the complex treatments, the second-stage model schedules the remaining treatments and performs the detailed allocation of patients to rooms on the wards respecting gender separation constraints.

Besides the fact that in the second-stage model, certain decisions are considered as fixed, the main differences between the first-stage and the second-stage model are the level of detail on which the bed resources are modeled and the fact that in the second-stage model, all treatment groups are scheduled under consideration of their resource requirements. As a result, assuming that in the second-stage model, the sets of treatment groups  $\mathcal{G}_p^{\text{res}}$  contains all relevant treatment groups instead of only the subset of complex key treatments, the constraint sets (1–4) dealing with scheduling ward stays (implying admission and discharge days) and (13–23) mainly dealing with treatment scheduling and the allocation treatment resources also appear in the second-stage model. In the following, we present the structurally different parts of the second-stage model, that is, the constraints for fixing the first-level decisions, the

model component for representing bed resources on a high level of detail and the objective function in which the bed-related terms make use of the variables introduced in the second-stage model.

### 4.3.1 Enforcing the instructions from the first-stage model

To enforce the instructions from the first-stage model in the second-stage model, we impose the following constraints fixing the respective variables: constraints set (25) fixes the admission day  $d_p^{ad}$  and the discharge day  $d_p^{dis}$  of each patient. Assuming that for each patient  $p$ , the set of treatment groups scheduled by the first-stage model is denoted with  $\mathcal{G}_p^{fixed}$  and that  $d_{p,g}^{treat}$  corresponds to the day on which treatment group  $g \in \mathcal{G}_p^{fixed}$  is scheduled, constraints (26) fix the respective scheduling decisions in the second-stage model:

$$X_{p,1}^{ws} = d_p^{ad}; X_{p,|S|+1}^{ws} - 1 = d_p^{dis} \quad \forall p \in \mathcal{P} \tag{25}$$

$$X_{p,g}^{treat} = d_{p,g}^{treat} \quad \forall p \in \mathcal{P}, g \in \mathcal{G}_p^{fixed} \tag{26}$$

### 4.3.2 Bed resources, room assignment and gender separation based on individual rooms

In contrast to the first-stage model in which on each ward, the rooms with a common bed capacity are aggregated, the second-stage model considers each room individually. For each day  $d$ , the binary day-indexed room stay variables  $Y_{p,s,q,d}^{room}$  determine whether a patient  $p$  stays in room  $q$  in ward stay  $s$ ;  $\mathcal{Q}_{p,s}$  denotes the set of rooms which can be used by patient  $p$  for ward stay  $s$ . In analogy to the constraints (5–7) in the first-stage model, constraints (27–29) link the  $Y_{p,s,q,d}^{room}$  variables to the ward stay start variables  $X_{p,s}^{ws}$ , enforcing the allocation of a bed in a room for each day of each ward stay:

$$\sum_{q \in \mathcal{Q}_{p,s}} \sum_{d \in \mathcal{D}} Y_{p,s,q,d}^{room} = U_{p,s}^{los} \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p \tag{27}$$

$$X_{p,s}^{ws} \leq d + M \left( 1 - \sum_{q \in \mathcal{Q}_{p,s}} Y_{p,s,q,d}^{room} \right) \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p, d \in \mathcal{D} \tag{28}$$

$$X_{p,s+1}^{ws} - 1 \geq d - M \left( 1 - \sum_{q \in \mathcal{Q}_{p,s}} Y_{p,s,q,d}^{room} \right) \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p, d \in \mathcal{D} \tag{29}$$

In the second-stage model, it is assumed that a patient resides in the same room during the full ward stay. The assignment of a room  $q$  to a patient  $p$  at her ward stay  $s$  is represented in the binary decision variable  $Z_{p,s,q}^{room}$ ; the constraints set (30) ensures that each patient is assigned exactly one room during each ward stay. Constraints (31) enforce that the values of the day-indexed room assignment variables  $Y_{p,s,q,d}^{room}$  are in line with the same-room requirement.

$$\sum_{q \in Q_{p,s}} Z_{p,s,q}^{\text{room}} = 1 \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p \tag{30}$$

$$Y_{p,s,q,d}^{\text{room}} \leq Z_{p,s,q}^{\text{room}} \quad \forall p \in \mathcal{P}, s \in \mathcal{S}_p, q \in Q_{p,s}, d \in \mathcal{D} \tag{31}$$

The gender separation requirement means that on each day in each room, all occupants share the same gender. In the model, the binary decision variable  $Y_{q,d}^{\text{mRoom}}$  is introduced;  $Y_{q,d}^{\text{mRoom}}$  equals 1 if room  $q$  is a “male room” and 0 if room  $q$  is a “female room” on day  $d$ . The constraints (32) and (33) enforce that for each day  $d$  and in each room  $q$ , the gender of the room occupants is consistent with the value of the variable  $Y_{q,d}^{\text{mRoom}}$  and at the same time ensure the room capacity is respected. It is assumed that the room capacity can be extended by extra beds registered by the variable  $U_{q,d}^{\text{mBed}+}$  for beds occupied by males and  $U_{q,d}^{\text{fBed}+}$  occupied by females. The set of constraints (34) ensures that for each room and for each day, the upper bound  $b_q^{\text{add}}$  for the number of extra beds is respected and that the gender of the extra bed occupants matches the gender assignment of the rooms on each day:

$$\sum_{p \in \mathcal{P}^m} \sum_{s \in \mathcal{S}_{p,q}} Y_{p,s,q,d}^{\text{room}} \leq b_q Y_{q,d}^{\text{mRoom}} + U_{q,d}^{\text{mBed}+} \quad \forall q \in \mathcal{Q}, d \in \mathcal{D} \tag{32}$$

$$\sum_{p \in \mathcal{P}^f} \sum_{s \in \mathcal{S}_{p,q}} Y_{p,s,q,d}^{\text{room}} \leq b_q (1 - Y_{q,d}^{\text{mRoom}}) + U_{q,d}^{\text{fBed}+} \quad \forall q \in \mathcal{Q}, d \in \mathcal{D} \tag{33}$$

$$U_{q,d}^{\text{mBed}+} \leq b_q^{\text{add}} Y_{q,d}^{\text{mRoom}}; U_{q,d}^{\text{fBed}+} \leq b_q^{\text{add}} (1 - Y_{q,d}^{\text{mRoom}}) \quad \forall q \in \mathcal{Q}, d \in \mathcal{D} \tag{34}$$

### 4.3.3 Objective function

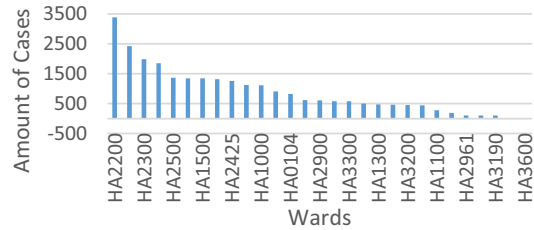
The objective function of the second-stage model is given by (35). While the scheduling objectives of both levels are basically identical, the objective function (35) reflects the main differences between the models on the two levels: first, since the admission decisions are fixed in the second stage, (35) does not consider admission day preferences. Second, since the level of detail of the bed allocation in the second-stage model is higher than in the first-stage model, the corresponding term is adapted:

$$\begin{aligned} \min & c^{\text{maxDel}} U^{\text{maxDel}} + \sum_{p \in \mathcal{P}} c^{\text{del}} U_p^{\text{del}} + \sum_{q \in \mathcal{Q}} \sum_{d \in \mathcal{D}} c^{\text{bed}+} (U_{q,d}^{\text{fBed}+} + U_{q,d}^{\text{mBed}+}) \\ & + \sum_{r \in \mathcal{R}} \left( c^{\text{maxOt}} U_r^{\text{maxOt}} + c^{\text{maxIt}} U_r^{\text{maxIt}} + \sum_{d \in \mathcal{D}} (c_d^{\text{ot}} U_{r,d}^{\text{ot}} + c_d^{\text{it}} U_{r,d}^{\text{it}}) \right) \end{aligned} \tag{35}$$

## 5 Real-world case study from a German university hospital

The research presented in this paper is based on a partnership with a German University Hospital providing us both with a real-world dataset and with an opportunity to discuss our assumptions and results with domain experts such as the

**Fig. 2** Distribution of elective patients on wards



case manager of the department of urology considered in our investigation. In 2011, the case study hospital had around 36,000 admissions divided into around 25,000 elective and 11,000 emergency inpatients. Figure 2 shows the amount of elective admissions on each ward in 2011. The department of urology (HA2200) has the highest number of cases; Fig. 3 shows the distribution of these cases grouped by month of admission.

We agreed with the partner hospital to base our analysis and our experiments on the peak month March 2011 exhibiting the highest number of cases. In close collaboration with the case manager, we prepared and validated the input data required for our approach including two sets objective function weights representing two different planning scenarios. The remainder of this section describes and illustrates these input data: Sect. 5.1 deals with the cases and their CP information automatically extracted by the pathway-mining approach sketched in Sect. 3. Section 5.2 describes how the data regarding resource requirements and resource capacities were obtained. Finally, Sect. 5.3 explains the planning scenarios used for the computational experiments.

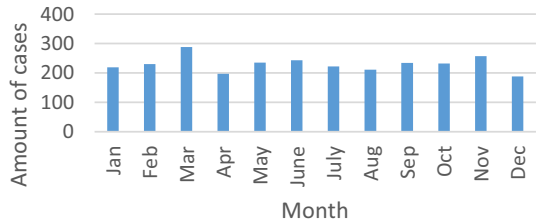
## 5.1 Case data and automated pathway extraction

As explained above, we conduct our investigation with the real-world data for cases admitted in March 2011 to the department of urology of a German university hospital. For these cases, we applied the CP mining approach proposed in Helbig et al. (2015) and shortly sketched in Sect. 3 of this paper. Since this mining approach operates with standardized hospital billing data according to the §21KHEntG<sup>2</sup> which has to be reported by each hospital in Germany, it is readily applicable for each German hospital. The pathway-mining approach identifies similar treatment profiles for each primary diagnosis and creates corresponding homogenous case groups. Based on these groups, feasible intervals for treatments and ward stays as well as order constraints are determined. Note that if only a single case with a certain treatment profile occurs in the database, the mining approach returns treatment chains without scheduling flexibility for treatments and ward stays.

In total, the data in the considered month encompass 286 cases (217 male and 69 female patients) with 90 different primary diagnoses and 69 different DRGs.

<sup>2</sup> More details about structure and containing data as well as an example dataset can be found at: [http://www.g-drg.de/cms/Datenveroeffentlichung\\_gem.\\_21\\_KHEntG](http://www.g-drg.de/cms/Datenveroeffentlichung_gem._21_KHEntG).

**Fig. 3** Distribution of cases on HA2200 in 2011



Applying the mining approach to the described dataset resulted in 229 different CP among which 136 are based on unique treatment profiles and 93 represent profiles occurring more than once. In total, the cases involve 1088 treatment groups and 302 ward stays; around 6% of all cases involve more than one ward stay.

Figures 4, 5, and 6 show different CP for the primary diagnosis benign neoplasm: rectum based on a single treatment profile (Fig. 4), prostatic hyperplasia, a profile occurring three times (Fig. 5) and the most frequent profile (Fig. 6) for the primary diagnosis of malignant tumors of the prostate. The CP depicted in Fig. 4 exhibits two ward stays; the first on the department of urology and the second on an ICU (after the ICU the patient is relocated to a rehabilitation). Figures 5 and 6 exhibit a single ward stay on the department of urology. Each CP contains a specific set of treatments based on the OPS-Codes (operations and procedures) documented in the billing data according to §21KHEntG. Figure 4 shows a unique CP without scheduling flexibility for the ward stays and treatment groups, Figs. 5 and 6 depict profiles occurring more than once exhibiting scheduling flexibility for the treatment groups.

## 5.2 Resource requirements and availability

In addition to the CP structure discussed so far, information regarding required and available resources is needed for our approach. To obtain a realistic set of parameters, we discussed all 259 different OPS-Codes that had occurred during our planning horizon (March 2011) with the case manager of the department for urology who provided us with information regarding the required resources and resource times for each code. Table 2 shows an excerpt of the required time on resource types needed by a treatment.

Since OPS-Codes are very detailed and surgeries usually involve multiple codes, considering each individual set of codes would typically result in a unique type of pathway for each surgery. To avoid this, all surgery codes (starting with a “5”) appearing at the same day of a CP are aggregated to a single artificial code. To obtain a robust approximation the resource requirements for such a surgery code, we consider the maximum time a single resource unit was required for any of the OPS-Codes aggregated to the surgery code.

This approximation is based on the assumption provided by the case manager that all required resources need to be available for the full surgery duration. This assumption is reflected by all of her resource requirement estimates for individual surgery OPS-Codes. Using the maximum duration over all involved OPS-Codes,

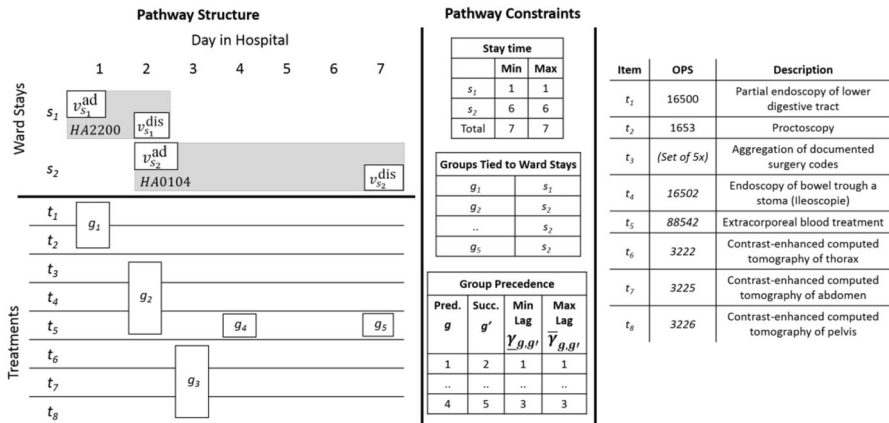


Fig. 4 Unique CP for benign neoplasm: rectum

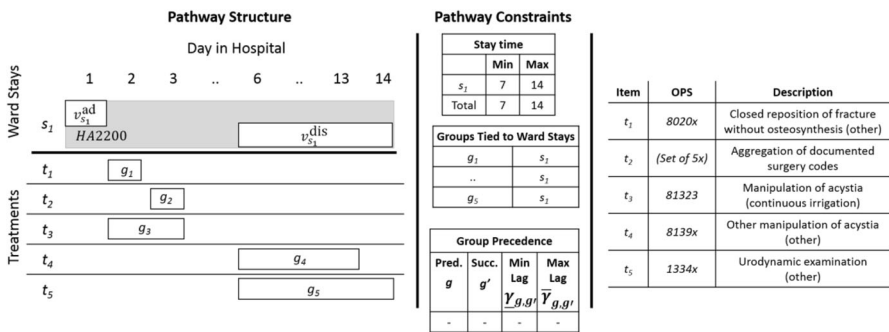


Fig. 5 CP for prostatic hyperplasia occurring 3 times

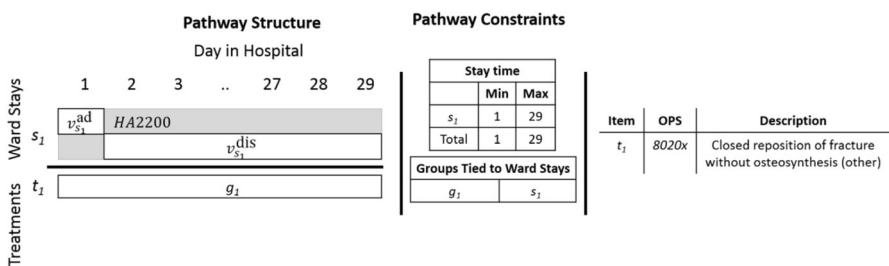


Fig. 6 CP for malignant tumors of the prostate occurring 75 times

however, tends to overestimate the actual surgery duration: Taking all cases from March 2011 considered in our study, on average the duration of a surgery is overestimated by 4.6 min. In accordance with the case manager from the involved department, we chose this slightly pessimistic estimation procedure for the surgery duration and resource consumption instead of a more optimistic estimate to make



**Table 2** Resource time required for OPS-Codes (excerpt)

OPS-code	Required time on resource type (h)								..
	Physician	Nurse	OR	OR-nurse	Deputy	Urography <sup>a</sup>	MRT	Anesthetist	
12750	0.5	0.5							..
13341	2	2				2			..
557840	2 × 0.75		0.75	0.75				0.75	..
560401	2 × 4		4	4	4			4	..
81372	0.5	1				0.5		0.5	..
..	..	..	..	..	..	..	..	..	..

This is a special kind of operating room located in the department of urology

the scheduling results more robust against uncertainty with respect to surgery durations.

Table 3 shows an excerpt of required resource types and their required time for some example surgeries; each surgery belongs to a primary diagnosis and a CP variant. For example, the surgery with the primary diagnosis C61 in variant 1 is based on the OPS-Codes 557,840 and 560,401. Using our assumptions, the required capacity for physicians and anesthetists is 8 h (2 physicians working 4 h) and 4 h.

The case study department has a capacity of 88 patients during workdays and 73 at weekends. It has 7 single-, 5 two-, 9 three- and 11 four-bed rooms. We assume that small rooms are not used for scheduling elective patients because of their flexibility in handling emergencies. Based on this assumption, we reduced the amount of available beds to 53 by keeping only the 11 four-bed rooms and 3 three-bed rooms. With regard to the CP to be scheduled, there are four additional departments in which beds are needed for ward stays. For these departments, we assume a bed capacity based on the observed number patient cases available in form of two-bed rooms.

Regarding the treatment resources, we consider scarce and special resources such as operating rooms and the urography individually. On the one hand, this accurately

**Table 3** Resource time required for surgeries (excerpt)

Diagnosis (ICD-code)	CP-variant	Required Resource Time (h)						.
		Physician	Deputy	OR-nurse	OR	Anesthetist		
N40	22	2	2	4	2	–	..	
N40	24	0.75	0.75	0.75	0.75	–	..	
C61	1	8	4	4	4	4	..	
..	..	..	..	..	..	..	..	

models resource eligibility for treatments requiring special equipment: as an example, the central OR is eligible for 139 treatments whereas the OR of the urology department is only eligible for 66 of these treatments. A more complex example arises in the Urography: There are ten treatments for which both the Uro-1 and the Uro-2 resource is eligible; however, there are 19 treatments for which only Uro-1 and 13 treatments for which only Uro-2 is eligible.

On the other hand, considering complex resources individually avoids disaggregation issues such as scheduling three 4-h surgeries on a day while there are only two OR available for 6 h on the same day. In addition, complex resources often exhibit different daily capacities since they may be shared among different departments. Other resources such as human resources with identical skills (e.g., OR-nurses) are aggregated.

In the considered planning month (March 2011), approximately 60% of all patients were elective. Based on this observation, we assume that 60% of the total available time of all treatment resources owned by the urology department can be scheduled for elective patients. Furthermore, the per-day availability of shared hospital resources such as central OR, MRT or computer tomography for the urology department was determined based on interviews with the department case manager. Table 4 shows an excerpt of the capacity available for scheduling elective patients per resource type for each weekday.

### 5.3 Scheduling scenarios considered in the computational experiments

The computational experiments conducted with the data described above are based on two main scenarios differing with respect to the resource allocation objectives. In the first scenario referred to as “smooth allocation”, the hospital aims at establishing a smooth resource allocation throughout the planning horizon. As noted by the case manager at our partner hospital department, a smooth resource allocation is favored by the staff members and establishes a similar level of

**Table 4** Resource capacity for elective patients (excerpt)

Resource type	Resource	Amount of available time (h)						
		Mon	Tue	Wed	Thu	Fri	Sat	Sun
Physician		58	58	58	58	58	15	10
OR	Central OR	10	10	10	6	10	0	0
	OR-HA2200	5	5	5	5	5	5	5
OR-nurse		24	24	24	24	24	5	5
Deputy		20	20	20	20	20	0	0
Urography	Uro-1	5	5	5	5	5	5	5
	Uro-2	10	10	10	10	10	10	10
	Uro-3	8	8	8	8	8	8	8
..	..	..	..	..	..	..	..	..

flexibility for reacting to emergency cases on each day of the planning horizon. To achieve a smooth allocation, the objective function coefficients associated with the min–max objectives as well as the penalty associated with daily resource overtime are both ten while the other objectives are smaller (see Table 5). Since extra beds on rooms are considered as worse than idle time or patient delays, we set extra bed penalty to five and the delay and idle time penalties to two.

In the second scenario referred to as “early allocation”, the hospital aims at establishing a high level of resource utilization at the beginning of the planning horizon. The scenario can be motivated by the observation from the case manager that the level of uncertainty affecting scheduling-relevant information regarding patient cases and resource availability increases towards the end of the planning horizon. This scenario is particularly useful in a setting where our approach is embedded in a rolling horizon approach. Establishing a high level of resource utilization at the beginning of the planning period aims at providing more flexibility for coping with the higher amount of uncertainty towards the end of the planning period. In addition, freeing future resource capacity can be instrumental for admitting a higher number of elective patients. The set of objective function coefficients for this scenario is depicted in the second row of Table 5; note that in particular, this scenario involves a quadratic decreasing day-dependent penalty for the idle time of resources and zero penalties for the maximum idle time and for the deviation from admission day penalties.

## 6 Experimental results

The real-world data described in the previous section forms the basis of a series of computational experiments with our hierarchical approach. The first set of experiments, discussed in Sect. 6.1, aims at assessing the hierarchical approach proposed in this paper. The second sets of experiments aim at discussing the results of the two scenarios in detail.

For the experiments, the MILP models presented in Sect. 4 were implemented AMPL; all treatments exhibiting a duration of more than 30 min (about 50% of the treatments) were considered to form the set  $\mathcal{G}_p^{\text{res}}$  of treatment groups to be scheduled in the first-stage model. The model instances were solved with Gurobi 6.5 on an Intel i7-3770 CPU with 3.40 GHz and 8.00 GB RAM using 6 threads. For solving the aggregated first-stage model, the relative MIP gap tolerance was set to 1%, the

**Table 5** Objective function weights in the two experimental scenarios

Scenario	$c^{\text{maxDel}}$	$c^{\text{del}}$	$c^{\text{ad}}$	$c^{\text{bed+}}$	$c^{\text{maxOt}}$	$c^{\text{maxIt}}$	$c_d^{\text{ot}}$	$c_d^{\text{it}}$
Smooth allocation	10	2	1	5	10	10	10	2
Early allocation	10	2	0	5	10	0	10	$5 + \frac{(31-d)^2}{100}$

time limit was set to 2 h and the MIP search strategy was set to focus on improving bounds. The second-stage model was solved using default Gurobi parameter settings.

### 6.1 Evaluating the hierarchical approach

Compared to approaches presented in the literature, our approach addresses a very detailed pathway-scheduling problem considering multiple ward stays, gender-separated room assignment and treatment-specific resource eligibility. In this section, we first compare the hierarchical two-stage approach to a monolithic MILP formulation. Next, we evaluate the effect of ignoring the gender separation requirement on the quality of the solution, and finally, we present experimental results showing the importance of carefully anticipating the second-stage problem in the first stage of our hierarchical solution approach.

#### 6.1.1 Monolithic model vs hierarchical approach

The main reason for introducing a hierarchical approach for the pathway-scheduling problem considered in this paper is that solving a monolithic MILP formulation of the problem with a standard solver is not possible in a reasonable amount of time for realistically sized problem instances. Table 6 depicts the results from experiments with both a monolithic model and the hierarchical approach proposed in Sect. 4 for the dataset described in Sect. 5. For the computation with the monolithic model, we allowed for a maximal solution time of 12 h and varied the maximum allowed deviation  $\bar{b}^{devAd}$  of the desired admission from 3 up to 10 days. The monolithic model instances exhibit 19,500 rows, 112,000 columns and 700,000 nonzeros; the instances of the first-stage model in the hierarchical approach contain 19,000 rows, 29,000 columns and 530,000 nonzeros; the instances of the second-stage model contain 2,200 rows, 6,200 columns and 25,000 nonzeros.

The results in Table 6 show that within the desired solution time of 2 h, the MILP solver was unable to find a feasible solution to any of the instances of the monolithic model. For instances permitting a higher deviation from the desired

**Table 6** Integrated vs. hierarchical approach

Scenario	$\bar{b}^{devAd}$	Monolithic				Hierarchical		Monolithic vs hierarchical Objective gap (%)
		Integer (h)	Total (h)	Gap (%)	Objective	Total (m)	Objective	
Smooth allocation	3	7.25	12	1.03	7856	5,2	8344	6.21
	5	11.6	12	4.14	8096	9,1	8299	2.50
	10	–	12	$\infty$		72,6	8101	–
Early allocation	3	3.5	3.9	1	25160	0,6	26,030	3.46
	5	–	12	$\infty$		3,2	25,549	–

admission day, this even holds after 12 h of solution time: it was only possible to find integer feasible solutions of allowed admission day deviations of up to 5 days in the smooth allocation scenario and for up to 3 days in the early allocation scenario. Furthermore, only for a single instance, a solution meeting the predefined 1% optimality gap is found within 12 h. For the hierarchical approach, the solution time is less than 10 min for all but one instance for which the solution time is less than 73 min.

The last column in Table 6 aims at evaluating the price at which this considerable decrease in solving time comes: For those instances for which a feasible solution could be found for the monolithic model, it depicts the percentage objective function gap between the solution obtained with the hierarchical approach and the best integer solution found for the monolithic model. It turns out that for this gap is between 2.5 and 6.2%; the highest gap is observed in the smooth allocation scenario with the smallest flexibility regarding the deviation from the desired admission day. The objective function gaps between the monolithic and the hierarchical approach can mainly be attributed to the fact that the first stage model determining the cornerstones of the CP schedule is based on an aggregate representation of room assignments and only considers the resource requirements for complex treatments. However, as will become clear in the next section, in contrast to the monolithic model, the hierarchical approach allows obtaining high-quality feasible solutions with a fully flexible admission date within a time limit of 2 h.

### 6.1.2 *The effect of ignoring gender separation*

Compared to other pathway-scheduling approaches, our model operates on a higher level of detail. With regard to treatments, our model considers the eligibility of resources based on resource properties (qualifications, available equipment) instead of characterizing resources using a single resource type (nurse, operating room) and stating requirements in terms of these resource types. While the advantage of this more realistic representation for matching treatments and resources cannot be evaluated experimentally in a straightforward way, in the following, we discuss results from comparing different levels of detail of representing the allocation of bed resources. To evaluate the relevance of considering gender-separated rooms, we compare the solution obtained with our approach (a priori consideration of gender separation) to a solution obtained by first scheduling pathways ignoring the gender separation requirement and afterwards determining a gender-separated room assignment is performed based on these schedules (a posteriori consideration of gender separation).

Table 7 shows the results from using this comparison for two instances, both based on the “early bed allocation” scenario, with an increased penalty of 50 for extra beds: in the first instance, the number of beds and rooms in the department of urology is as described in Sect. 5 (11 four-bed rooms and 3 three-bed rooms) while in the second instance, only 10 four-bed rooms are considered. It turns out that in both instances, the overall solution quality is better and the number of needed extra beds is smaller when considering gender separation a priori. While the objective

**Table 7** Amount of additional beds respecting and ignoring gender restrictions

Instance	Considering gender a priori		Considering gender a posteriori			
	A priori		A priori		A posteriori	
	Extra beds	Objective	Extra beds	Objective	Extra beds	Objective
Given capacity	0	24,284	1	24,194	3	24,294
Reduced capacity	19	25,414	16	25,181	28	25,781

function differences are comparably small, note that a significant amount of extra beds typically represents solutions which are not implementable in practice.

### 6.1.3 The role of anticipation in the hierarchical approach

In our hierarchical approach, we carefully designed the anticipation of the second-stage problem in the first stage. As an example, we propose to consider all treatment groups (not only those to be scheduled in the first-stage model) to anticipate the full set of precedence constraints in the first-stage model. In our experiments, this anticipation turned out to be crucial: without this anticipation, almost all the second-stage model instances were infeasible. As depicted in Table 8 showing results from based on the “early allocation” scenario with the reduced room capacity setting described above, it is also important to anticipate the gender separation requirement for patient rooms: if the gender separation anticipation is dropped from the first-stage model, the final solution obtained from the second-stage model involves considerably more extra beds.

## 6.2 Smooth resource allocation scenario: detailed results

In the first scenario to be discussed in detail, the objective structure favors a smooth allocation of resource utilization over the planning horizon. Table 9 shows the solution times and objective function values for different settings of the parameter  $\bar{b}^{devAd}$  limiting the deviation from the preferred admission day from 0 to 30. The results show that the solution time of the first-stage model is very sensitive to the flexibility regarding the admission day: The model with a flexibility of 5 days is

**Table 8** Comparing the amount of extra beds

$\bar{b}^{devAd}$	With gender anticipation		Without gender anticipation	
	Extra beds	Objective value	Extra beds	Objective value
10	26	26,019	37	26,577
30	19	25,414	29	26,053

**Table 9** Solving time, gaps and objectives in the smooth resource allocation scenario

$\bar{b}^{\text{devAd}}$	Solution time 1st model (s)	Gap (%)	Objective	Solving time 2nd model (s)	Gap (%)	Objective
0	4	1	9228	20	0	9066
3	310	1	8493	1	0	8212
5	543	1	8435	1	0	8150
10	4352	1	8264	1	0	7919
30	3069	1	8253	1	0	7898

solved in less than 10 min whereas the model with 10 days needs an eight times higher solution time. Surprisingly, the solution time of the last instance (30 days) is shorter than for the 10-day instance. We suspect that that this is due the fact that the increased admission day flexibility avoids resource bottlenecks. The second-stage model could easily be solved in less than 20 s in all cases but the first.

Table 10 provides more details regarding the effects of increasing admission day flexibility: Even a small amount of flexibility considerably improves resource- and patient-related performance indicators. The total amount of overtime for OR, OR-nurses and urography is reduced by around 92% in case of allowing the admission day to deviate from the desired day by 3 days and up to 96% in the 30-day case. Considering the OR values in detail (two central OR and one ward-owned OR) shows that the total overtime in the 30-day case is allocated to both central OR with a maximum overtime of 0.7 h (the maximum overtime in the zero-days-case was 6 h). The ward-owned OR which cannot perform all kinds of surgeries needs no additional overtime at all in the 30-day case anymore.

With an increasing flexibility, the idle times of the OR are highly reduced. This contrasts with the comparably small change of idle time for OR-nurses and urography—we suspect that these resources are not scarce. Taking a closer look, we can confirm this observation for OR-nurses and for one of the urography resources. The other urography, however, forms a scarce resource since it is eligible for a wider range of treatments. With regard to patient-related indicators, the maximum delay is reduced from 3 to 0 days and the total amount of all delays decreases from 46 to 0 days. On the other hand, the distance to the preferred admission day increases. Figure 7 depicts the frequencies of the deviations from the desired admission days for all patients for the last instance: from all 286 patients, 231 can be admitted at their preferred day; 7% of all patients are admitted before and 11% are admitted after their preferred admission day.

Figure 8 visualizes the allocation profiles of the resource type physician, operating room and urography as given by empirical ex-post data (left hand side) and after the scheduling with an admission day flexibility of 30 days (right hand side). The blue background depicts the available capacity of treatment time and the visible area corresponds to the remaining capacity. The black line on top is the goal

**Table 10** Key indicators in the smooth resource allocation scenario

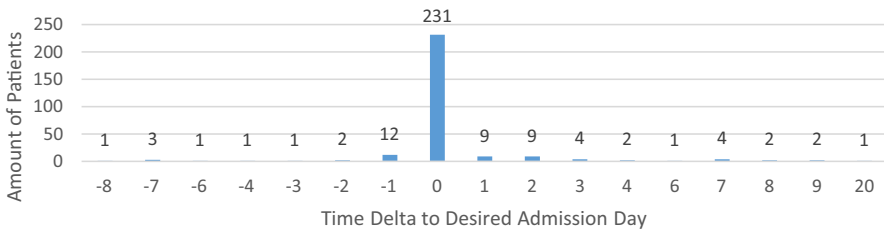
Indicator	$\bar{b}^{devAd}$			
	0	3	10	30
Maximum patient delay	3	5	0	0
Total patient delay	46	23	0	0
Maximum overtime				
OR	6	1.8	0.7	0.7
OR-nurse	2	0	0	0
Uro	3.5	0.75	0.2	0
Maximum idle time				
OR	12	7.8	4.2	4.7
OR-nurse	19.2	10.2	8.7	7.7
Uro	11.5	10.3	10.5	9.5
Total overtime				
OR	36.8	2	0.85	1.75
OR-nurse	3.6	0	0	0
Uro	4.2	1.25	0.23	0
Total idle time				
OP	74.9	40.5	39	39.9
OR-nurse	135.7	132.8	132	132
Uro	192.7	189.9	188.8	188.6
Extra male beds	1	0	0	0
Extra female beds	1	0	0	0

ratio of 80% capacity allocation. The red and the green areas are the allocated resource times at each day.

For each resource, the allocation obtained by our approach is very smooth compared to the empirical data. The resource allocation of the physicians shows an even workload distribution. Given the large blue area, considering only elective patients, physicians and OR-nurses obviously do not form scarce resources. Regarding the combined allocation profiles of all operating rooms, the results obtained by our approach exhibit a better fit to the available capacity profile with less overtime occurring during the month compared to the empirical data. Since there is no idle time for the central OR, the visible available capacity can be attributed exclusively to the ward-owned OR.

To quantify the smoothing effect obtained by our scheduling approach, let us consider the standard deviation of the daily allocation of each resource: For physicians, the standard deviation change from 12.7 to 7.1, for operating rooms from 7 to 4, for OR-nurses from 8.4 to 5.2 and for the urography from 4.6 to 2.2. Given these values, it can be summarized that the resource allocation is successfully smoothed in this scenario. Furthermore, even under consideration of gender separation, no extra beds are required to hospitalize all 286 elective patients.





**Fig. 7** Time delta between preferred and scheduled admission—equal allocation

### 6.3 Early resource allocation scenario: detailed results

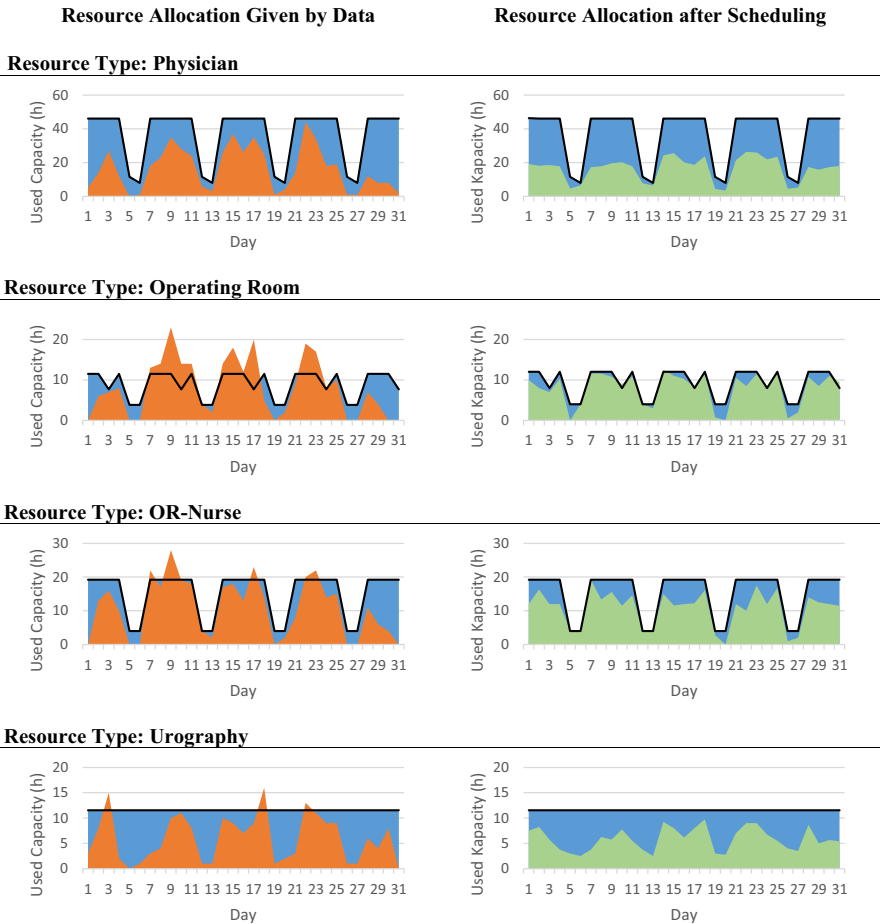
While the scenario presented in the previous section aimed at obtaining a smooth allocation of resources, the present section focuses on a scenario in which scheduling favors a high resource utilization at the beginning of the planning horizon. As in the first scenario, we vary the maximum allowed deviation from the preferred admission day by changing the parameter

$\bar{b}^{\text{devAd}}$  from 0 to 30 days for the first-stage model. Table 11 shows the objective function values, solving times and gaps for each considered value of  $\bar{b}^{\text{devAd}}$ . It turns out that even the largest possible instance with completely flexible patient admission days was solved in less than 2 h. We suspect that this considerable decrease in solving time compared to the first scenario results from the fact that in the second scenario, the objective function is less symmetric. For all variants, the second model is solved to optimality in less than 10 min.

Table 12 gives an overview on how selected indicators are affected by increasing admission date flexibility from 0 up to 30 days. The patient-related indicators (maximum and individual delay) can be reduced to zero. The resource-related indicators, however, exhibit a more complex pattern: while the total over- and idle time is reduced from the 0-day to the 10-day case, these values increase in the 30-day case. The reason for this behavior is patients shifted from the end of the scheduling horizon to the beginning to avoid most expensive idle times on all resources.

Figure 9 shows the absolute frequencies of the deviations between scheduled and the preferred admission days for the case of a fully flexible admission day. The scheduled admission days deviate up to 28 days from the preferred admission day. 65% (186) of the patients are admitted earlier than preferred; 30% (80 patients) are admitted more than 10 days earlier. On average, each patient is 4.5 days earlier admitted.

The effect of the early resource allocation is visualized in Fig. 10 showing the time-dependent resource allocation profile of physicians, OR, OR-nurses and urography based on historical ex-post data (left hand side) as well as based on scheduling with full admission day flexibility (right hand side). The results show that the goal of establishing a high resource utilization at the beginning of the planning horizon could be achieved.



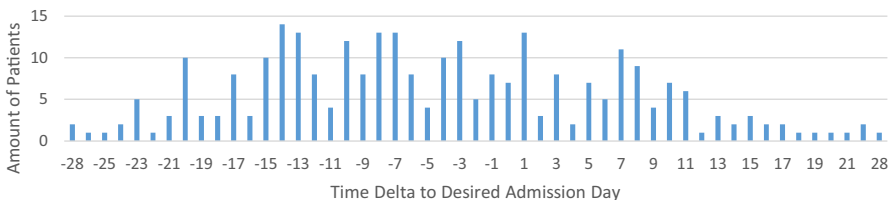
**Fig. 8** Results of smooth resource allocation with 10-day deviation

**Table 11** Solving time, gaps and objectives aiming on early high resource allocation

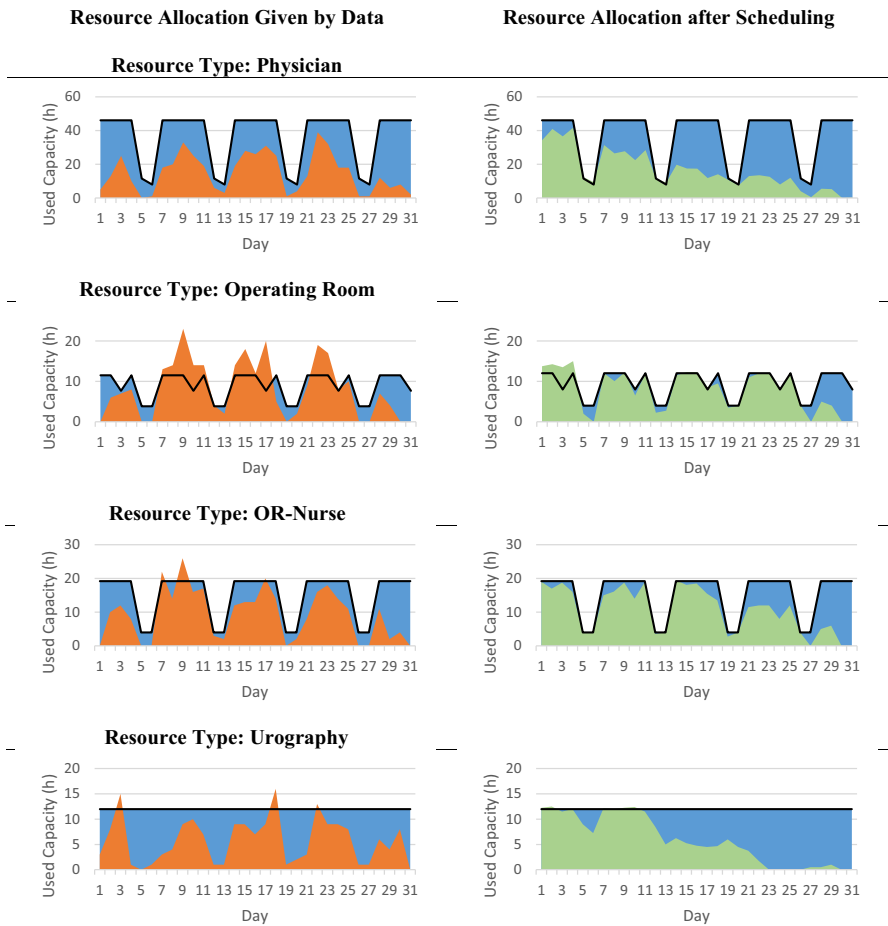
$\bar{b}^{devAd}$	Solving time 1st model (s)	Gap (%)	Objective	Solving time 2nd model (s)	Gap (%)	Objective
0	1	1	27,982	13	0	27,384
3	34	1	26,490	1	0	26,030
5	190	1	26,062	4	0	25,549
10	870	1	25,160	24	0	24,679
30	4488	1	24,726	506	0	24,186

**Table 12** Key indicators in the high early allocation scenario

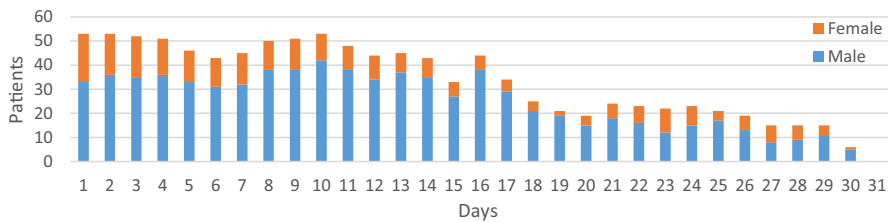
Indicators	$\bar{b}^{devAd}$			
	0	3	10	30
Maximum patient delay	6	1	0	0
Total patient delay	45	2	0	0
Maximum overtime				
OP	7	3.8	2.4	4.7
OR-nurse	1.3	0.4	0.1	0.3
Uro	3.5	0.2	0.4	0.7
Total overtime				
OP	39.1	13.1	8.6	14.8
OR-nurse	3.6	0.8	0.1	0.3
Uro	4.5	0.3	0.2	0.3
Total idle time				
OP	77.2	51.3	46.9	52.9
OR-nurse	135.7	132.8	132.2	132.4
Uro	193.2	189	189.6	190
Extra male beds	1	0	4	2
Extra female beds	1	3	1	0

**Fig. 9** Deviations of scheduled from preferred admission day—early allocation

Since at the beginning of the planning horizon, the penalty for idle time substantially outweighs the penalties for overtime, the scarce OR resources tend to exhibit overtime while other resources are highly utilized. Note that an even higher utilization is prevented by the scarcity of bed resources at the admission ward, see Fig. 11 for a visualization of the bed allocation over time. At the first weekend starting at day 5, the OR exhibits a high amount of idle time: in this case, physicians and OR-nurses form the scarce resources, and reducing OR idle time would induce overtime for several resources at once. Interestingly, the scheduling results on the first weekend also reflect the fact that in contrast to OR, the urography resources do not require the presence of OR-nurses: their utilization does not decrease as heavily as the OR utilization; furthermore, given the reduced physician availability on Sundays, the urography utilization is lowest on Sunday.



**Fig. 10** Results of early resource allocation with 30-day deviation



**Fig. 11** Patient assignment

Summarizing the results of this scenario, the desired maximization of the resource allocation at the beginning of the planning horizon can be achieved. However, putting the described results into practice may turn out to be risky since in

case of a rate of arriving emergency patients higher than the expected ratio of 40%, capacity bottlenecks will arise at least during the first week. Nevertheless, our scheduling approach can be used to identify such bottlenecks and to take measures to avoid them or to mitigate their consequences. In the current case, operating rooms and OR-nurses seem to be the scarcest resource types. This becomes particularly obvious after scheduling, showing noticeable available capacity only during the last days of the planning horizon.

## 7 Conclusion

This paper presents a hierarchical MILP approach for CP scheduling aiming at practical applicability in several ways. The CP information including the pathway constraints can be automatically extracted from standardized hospital billing data. The data-driven character of our approach substantially reduces the amount of work to be performed to obtain the CP information needed for scheduling. Moreover, the CP scheduling approach considers all relevant resources on an adequate level of detail; in particular, it models bed assignment on the level of rooms and allows considering treatment resources individually to avoid disaggregation problems of complex resources such as operating rooms. Furthermore, it considers several practically relevant aspects such as multiple ward stays per hospitalization, gender-separated rooms and eligibility of resources for treatments (e.g., special equipment in operating rooms needed for a certain surgery or special qualifications of nurses). To support a broad range of application scenarios and hospital-specific objective structures, our approach employs a multi-criteria objective function accounting for both patient- and hospital-related objectives. Since the detailed representation of practically relevant aspects yields complex large-scale model instances, we propose a carefully designed hierarchical two-stage approach involving anticipation components aligning the first-stage solution with aspects relevant to the second-stage problem.

The experimental results based a case study with real-world data involving a planning horizon of 30 days, 1088 treatments and 302 ward stays of 286 patients show that our approach is capable of obtaining high-quality solutions within a reasonable amount of time: even the most complex instances in which the admission days of the patients can be freely chosen can be solved within less than 90 min. Furthermore, the detailed evaluation shows that our model is capable of considering different scenarios such as achieving a smooth allocation of resources over the planning horizon and establishing a high resource allocation in the beginning of the horizon.

The results presented in this paper offer multiple opportunities for future research: The robustness of the schedules obtained with our approach could be assessed by a simulation study accounting for sources of uncertainty such as such as emergency patients, stochastic treatment times and short-term resource unavailability. Moreover, evaluating our approach in additional settings would be instrumental for understanding in how far the results presented in this paper can be generalized. From a practical perspective, the main goal of our efforts is to

embed our model into the information systems of our partner university hospital to provide decision support for hospital-wide scheduling of clinical pathways. While we are aware that this goal is presently out of reach mainly due to the present organizational structure in hospitals and due to a lack of integrated information systems, we believe that providing an approach that is data-driven and considers many practically relevant aspects is a step in the right direction.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Artigues, Christian, Oumar Koné, Pierre Lopez, and Marcel Mongeau. 2015. Mixed-Integer Linear Programming Formulations. *Handbook on Project Management and Scheduling. International Handbooks on Information Systems*, vol. 1, 17–41. Cham: Springer. [https://doi.org/10.1007/978-3-319-05443-8\\_2](https://doi.org/10.1007/978-3-319-05443-8_2).
- Augurzyk, Boris (ed.). 2015. *Bad Bank für Krankenhäuser: Krankenhausausstieg vor der Tür?; [inklusive eBook] Krankenhaus-Rating-Report*. Essen: HCB, Institute for Health Care Business GmbH; RWI.
- Augusto, Vincent, Xiaolan Xie, and Viviana Perdomo. 2010. Operating theatre scheduling with patient recovery in both operating rooms and recovery beds. *Computers & Industrial Engineering* 58: 231–238. <https://doi.org/10.1016/j.cie.2009.04.019>.
- Banditori, C., P. Cappanera, and F. Visintin. 2013. A combined optimization-simulation approach to the master surgical scheduling problem. *IMA Journal of Management Mathematics* 24: 155–187. <https://doi.org/10.1093/imaman/dps033>.
- Batun, Sakine, Brian T. Denton, Todd R. Huschka, and Andrew J. Schaefer. 2011. Operating Room Pooling and Parallel Surgery Processing Under Uncertainty. *INFORMS Journal on Computing* 23: 220–237. <https://doi.org/10.1287/ijoc.1100.0396>.
- Burdett, Robert L., and Erhan Kozan. 2018. An integrated approach for scheduling health care activities in a hospital. *European Journal of Operational Research* 264: 756–773. <https://doi.org/10.1016/j.ejor.2017.06.051>.
- Cardoen, Brecht, and Erik Demeulemeester. 2011. A Decision Support System For Surgery Sequencing at UZ Leuven's Day-Care Department. *International Journal of Information Technology & Decision Making* 10: 435–450. <https://doi.org/10.1142/S0219622011004397>.
- Cardoen, Brecht, Erik Demeulemeester, and Jeroen Beliën. 2009a. Optimizing a multiple objective surgical case sequencing problem. *International Journal of Production Economics* 119: 354–366.
- Cardoen, Brecht, Erik Demeulemeester, and Jeroen Beliën. 2009b. Sequencing surgical cases in a day-care environment: an exact branch-and-price approach. *Computers & Operations Research* 36: 2660–2669. <https://doi.org/10.1016/j.cor.2008.11.012>.
- Ceschia, Sara, and Andrea Schaerf. 2011. Local search and lower bounds for the patient admission scheduling problem. *Computers & Operations Research* 38: 1452–1463. <https://doi.org/10.1016/j.cor.2011.01.007>.
- Ceschia, Sara, and Andrea Schaerf. 2014. Dynamic patient admission scheduling with operating room constraints, flexible horizons, and patient delays. *Journal of Scheduling*. <https://doi.org/10.1007/s10951-014-0407-8>.
- Chaabane, Sondes, Nadine Meskens, Alain Guinet, and Marius Laurent. 2008. Comparison of two methods of operating theatre planning: application in Belgian Hospital. *Journal of Systems Science and Systems Engineering* 17: 171–186. <https://doi.org/10.1007/s11518-008-5074-x>.
- Chow, Vincent S., Martin L. Puterman, Neda Salehirad, Wenhui Huang, and Derek Atkins. 2011. Reducing Surgical Ward Congestion Through Improved Surgical Scheduling and Uncapacitated Simulation: reducing Surgical Ward Congestion. *Production and Operations Management* 20: 418–430. <https://doi.org/10.1111/j.1937-5956.2011.01226.x>.

- Clavel, Daniel, Cristian Mahulea, Jorge Albareda, and Manuel Silva. 2017. Towards efficient algorithms for planning surgeries in operation rooms. In *2017 22nd IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 1–8. Limassol: IEEE. <https://doi.org/10.1109/etfa.2017.8247620>.
- Conforti, Domenico, Francesca Guerriero, Rosita Guido, Marco Matucci Cerinic, and Maria Letizia Conforti. 2011. An optimal decision making model for supporting week hospital management. *Health Care Management Science* 14: 74–88. <https://doi.org/10.1007/s10729-010-9144-5>.
- Correia, Isabel, and Francisco Saldanha-da-Gama. 2015. A Modeling Framework for Project Staffing and Scheduling Problems. In *Handbook on Project Management and Scheduling Vol.1*, 547–564. International Handbooks on Information Systems. Springer, Cham. [https://doi.org/10.1007/978-3-319-05443-8\\_25](https://doi.org/10.1007/978-3-319-05443-8_25).
- Demeester, Peter, Wouter Souffriau, Patrick De Causmaecker, and Greet Vanden Berghe. 2010. A hybrid tabu search algorithm for automatically assigning patients to beds. *Artificial Intelligence in Medicine* 48: 61–70. <https://doi.org/10.1016/j.artmed.2009.09.001>.
- Denton, Brian T., Andrew J. Miller, Hari J. Balasubramanian, and Todd R. Huschka. 2010. Optimal Allocation of Surgery Blocks to Operating Rooms Under Uncertainty. *Operations Research* 58: 802–816. <https://doi.org/10.1287/opre.1090.0791>.
- Denton, Brian T., James Viapiano, and Andrea Vogl. 2007. Optimization of surgery sequencing and scheduling decisions under uncertainty. *Health Care Management Science* 10: 13–24. <https://doi.org/10.1007/s10729-006-9005-4>.
- Fei, H., C. Chu, and N. Meskens. 2009. Solving a tactical operating room planning problem by a column-generation-based heuristic procedure with four criteria. *Annals of Operations Research* 166: 91–108. <https://doi.org/10.1007/s10479-008-0413-3>.
- Fei, H., N. Meskens, and C. Chu. 2010. A planning and scheduling problem for an operating theatre using an open scheduling strategy. *Computers & Industrial Engineering* 58: 221–230. <https://doi.org/10.1016/j.cie.2009.02.012>.
- Gartner, Daniel, and Rainer Kolisch. 2014. Scheduling the hospital-wide flow of elective patients. *European Journal of Operational Research* 233: 689–699. <https://doi.org/10.1016/j.ejor.2013.08.026>.
- Helbig, Karsten. 2011. Zeitplanung für Patientenpfade unter Berücksichtigung von Betten-, Behandlungskapazitäten und Fairnesskriterien. *Tagungsband 15. Interuniversitäres Doktorandenseminar Wirtschaftsinformatik der Universitäten Chemnitz, Dresden, Freiberg, Halle-Wittenberg, Jena und Leipzig*: 34–44.
- Helbig, Karsten, Michael Römer, and Taïeb Mellouli. 2015. A Clinical Pathway Mining Approach to Enable Scheduling of Hospital Relocations and Treatment Services. In *Business Process Management*, ed. Hamid Reza Motahari-Nezhad, Jan Recker, and Matthias Weidlich, 9253:242–250. Cham: Springer International Publishing.
- Helm, Jonathan E., and Mark P. Van Oyen. 2014. Design and Optimization Methods for Elective Hospital Admissions. *Operations Research* 62: 1265–1282. <https://doi.org/10.1287/opre.2014.1317>.
- Hulshof, Peter J.H., Nikky Kortbeek, Richard J. Boucherie, Erwin W. Hans, and Piet J.M. Bakker. 2012. Taxonomic classification of planning decisions in health care: a structured review of the state of the art in OR/MS. *Health Systems* 1: 129–175. <https://doi.org/10.1057/hs.2012.18>.
- Jacobs, Birgit. 2007. Ableitung von klinischen Pfaden aus evidenzbasierten Leitlinien am Beispiel der Behandlung des Mammakarzinoms der Frau. Universität Duisburg-Essen, Medizinische Fakultät/ Universitätsklinikum » Universitätsklinikum Essen » Institut für Medizinische Informatik, Biometrie und Epidemiologie.
- Jebali, Aïda, Atidel B. Hadj Alouane, and Pierre Ladet. 2006. Operating rooms scheduling. *International Journal of Production Economics* 99: 52–62. <https://doi.org/10.1016/j.ijpe.2004.12.006>.
- Kirschner, S., W.-C. Witzleb, M. Eberlein-Gonska, F. Krummenauer, and K.-P. Günther. 2007. Klinische Pfade. *Der Orthopäde* 36: 516–522. <https://doi.org/10.1007/s00132-007-1098-z>.
- Küttner, T., and N. Roeder. 2007. Definition Klinischer Behandlungspfade. *Klinische Behandlungspfade: mit Standards erfolgreich arbeiten/N. Roeder/T. Küttner (Hrsg.)–Köln: Dr. Ärzte-Verl., S: 19–27*.
- Lamiri, Mehdi, Johann Dreö, and Xiaolan Xie. 2007. Operating Room Planning with Random Surgery Times. In, 521–526. IEEE. <https://doi.org/10.1109/coase.2007.4341749>.
- Lamiri, Mehdi, Xiaolan Xie, Alexandre Dolgui, and Frédéric Grimaud. 2008a. A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research* 185: 1026–1037. <https://doi.org/10.1016/j.ejor.2006.02.057>.

- Lamiri, Mehdi, Xiaolan Xie, and Shuguang Zhang. 2008b. Column generation approach to operating theater planning with elective and emergency patients. *IIE Transactions* 40: 838–852. <https://doi.org/10.1080/07408170802165831>.
- Li, Xiangyong, N. Rafaliya, M. Fazle Baki, and Ben A. Chaouch. 2015. Scheduling elective surgeries: the tradeoff among bed capacity, waiting patients and operating room utilization using goal programming. *Health Care Management Science*. <https://doi.org/10.1007/s10729-015-9334-2>.
- Marques, Inês, M. Eugénia Captivo, and Margarida Vaz Pato. 2012. An integer programming approach to elective surgery scheduling: analysis and comparison based on a real case. *OR Spectrum* 34: 407–427. <https://doi.org/10.1007/s00291-011-0279-7>.
- Meskens, Nadine, David Duvivier, and Arnauld Hanset. 2013. Multi-objective operating room scheduling considering desiderata of the surgical team. *Decision Support Systems* 55: 650–659. <https://doi.org/10.1016/j.dss.2012.10.019>.
- Pham, Dinh-Nguyen, and Andreas Klinkert. 2008. Surgical case scheduling as a generalized job shop scheduling problem. *European Journal of Operational Research* 185: 1011–1025.
- Destatis. 2014. Publikation - Gesundheit - Statistisches Bundesamt (Destatis) - Gesundheit - Ausgaben - Fachserie 12 Reihe 7.1.1 - 2011 - Statistisches Bundesamt. [https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Gesundheitsausgaben/AusgabenGesundheitPDF\\_2120711.html](https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Gesundheitsausgaben/AusgabenGesundheitPDF_2120711.html). Accessed March 21.
- Riise, Atle, and Edmund K. Burke. 2011. Local search for the surgery admission planning problem. *Journal of Heuristics* 17: 389–414. <https://doi.org/10.1007/s10732-010-9139-x>.
- Roeder, N., P. Hensen, D. Hindle, N. Loskamp, and H.-J. Lakomek. 2003. Instrumente zur Behandlungsoptimierung. *Der Chirurg* 74: 1149–1155. <https://doi.org/10.1007/s00104-003-0754-z>.
- Rojas, Eric, Jorge Munoz-Gama, Marcos Sepúlveda, and Daniel Capurro. 2016. Process mining in healthcare: a literature review. *Journal of Biomedical Informatics* 61: 224–236. <https://doi.org/10.1016/j.jbi.2016.04.007>.
- Roland, B., C. Di Martinelly, F. Riane, and Y. Pochet. 2010. Scheduling an operating theatre under human resource constraints. *Computers & Industrial Engineering* 58: 212–220. <https://doi.org/10.1016/j.cie.2009.01.005>.
- Saadani, Nour El Houda, Zied Bahroun, and Asma Bouras. 2014. A linear mathematical model for patients' activities scheduling on hospital resources. In, 074–080. IEEE. <https://doi.org/10.1109/codit.2014.6996871>.
- Salfeld, Rainer, S. Hehner, Steffen Hehner, and Reinhard Wichels. 2009. *Modernes Krankenhausmanagement: Konzepte und Lösungen*. Springer.
- Samudra, Michael, Carla Van Riet, Erik Demeulemeester, Brecht Cardoen, Nancy Vansteenkiste, and Frank E. Rademakers. 2016. Scheduling operating rooms: achievements, challenges and pitfalls. *Journal of Scheduling* 19: 493–525. <https://doi.org/10.1007/s10951-016-0489-6>.
- Schimmelpfeng, Katja, Stefan Helber, and Steffen Kasper. 2012. Decision support for rehabilitation hospital scheduling. *OR Spectrum* 34: 461–489. <https://doi.org/10.1007/s00291-011-0273-0>.
- Schmidt, Robert, Sandra Geisler, and Cord Spreckelsen. 2013. Decision support for hospital bed management using adaptable individual length of stay estimations and shared resources. *BMC Medical Informatics and Decision Making* 13: 3.
- Schneeweiss, Ch. 1998. Hierarchical planning in organizations: elements of a general theory. *International Journal of Production Economics* 56–57: 547–556. [https://doi.org/10.1016/S0925-5273\(96\)00109-0](https://doi.org/10.1016/S0925-5273(96)00109-0).
- Sun, Bofei, Jie Song, and Leyuan Shi. 2013. Integrated operating room scheduling optimization under particularly workload constrain of surgeon. In, 972–977. IEEE. <https://doi.org/10.1109/coase.2013.6653965>.
- Testi, Angela, and Elena Tanfani. 2009. Tactical and operational decisions for operating room planning: efficiency and welfare implications. *Health Care Management Science* 12: 363–373. <https://doi.org/10.1007/s10729-008-9093-4>.
- Vancroonenburg, Wim, Patrick De Causmaecker, Frits Spieksma, and Greet Vanden Berghe. 2013. Scheduling elective patient admissions considering room assignment and operating theatre capacity constraints. In *Proceedings of the 5th International Conference on Applied Operational Research, Lecture Notes in Management Science*, 5:153–158.
- Vancroonenburg, Wim, Patrick De Causmaecker, Frits Spieksma, and Greet Vanden Berghe. 2014. Integrating patient-to-room assignment planning with elective admission scheduling. In *Proceedings of the 28th Annual Conference of the Belgian Operations Research Society*.



- Vanhaecht, K., M. Bollmann, K. Bower, C. Gallagher, A. Gardini, J. Guezo, U. Jansen, et al. 2006. Prevalence and use of clinical pathways in 23 countries - an international survey by the European Pathway Association. *International Journal of Care Pathways* 10: 28–34. <https://doi.org/10.1258/j.jicp.2006.124>.
- Villa, Stefano, Marta Barbieri, and Federico Lega. 2009. Restructuring patient flow logistics around patient care needs: implications and practicalities from three critical cases. *Health Care Management Science* 12: 155–165.
- Vissers, J.M.H. 2005. Patient mix optimization in tactical cardiothoracic surgery planning: a case study. *IMA Journal of Management Mathematics* 16: 281–304. <https://doi.org/10.1093/imaman/dpi023>.
- Vlah Jerić, Silvija, and José Rui Figueira. 2010. An Interactive Approach for Multiple Criteria Scheduling in a Croatian Hospital. *Croatian Operational Research Review* 1: 93–103.
- Vlah Jerić, Silvija, and José Rui Figueira. 2012. Multi-objective scheduling and a resource allocation problem in hospitals. *Journal of Scheduling* 15: 513–535. <https://doi.org/10.1007/s10951-012-0278-9>.
- Yang, W., and Q. Su. 2014. Process mining for clinical pathway: Literature review and future directions. In *2014 11th International Conference on Service Systems and Service Management (ICSSSM)*, 1–5. <https://doi.org/10.1109/icsssm.2014.6943412>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.