Hospital-wide Therapist Scheduling and Routing: Exact and Heuristic Methods

Daniel Gartner[1], Markus Frey[2] and Rainer Kolisch[2,3]

[1] School of Mathematics, Cardiff University, United Kingdom, gartnerd@cardiff.ac.uk

[2] TUM School of Management, Technische Universität München, Germany, markus.frey@tum.de, rainer.kolisch@tum.de

[3] Edward P. Fitts Department of Industrial and Systems Engineering, North Carolina State University, USA

Initial submission: March 17, 2016

Revision 1: February 11, 2017

Revision 2: June 22, 2017

Revision 3: December 14, 2017

Revision 4: September 17, 2018

Hospital-wide Therapist Scheduling and Routing: Exact and Heuristic Methods

Abstract

In this paper, we address the problem of scheduling and routing physical therapists hospital-wide. At the beginning of a day, therapy jobs are known to a hospital's physical therapy scheduler who decides for each therapy job when, where and by which therapist a job is performed. If a therapist is assigned to a sequence which contains two consecutive jobs that must take place in different treatment rooms, then transfer times must be considered. We propose three approaches to solve the problem. First, an Integer Program (IP) simultaneously schedules therapies and routes therapists. Second, a cutting plane algorithm iteratively solves the therapy scheduling problem without routing constraints and adds cuts to exclude schedules which have no feasible routes. Since hospitals are interested in obtaining quick solutions, we also propose a heuristic algorithm, which schedules therapies sequentially by simultaneously checking routing and resource constraints. Using real-world data from a hospital, we compare the performance of the three approaches. Our computational analysis reveals that our IP formulation fails to solve test instances, which have more than 30 jobs, to optimality in an acceptable solution time. In contrast, the cutting plane algorithm can solve instances with more than 100 jobs optimally. The heuristic approach obtains good solutions for large real-world instances within fractions of a second.

Keywords: Therapist scheduling; Job shop scheduling; Integer programming; Heuristics

1 Introduction

The demand for physical therapy services in hospitals is growing as aging baby boomers have increasing co-morbidities which require inpatient physical therapy (Bureau of Labor Statistics (2012), Vetrano et al. (2014)). In this paper, we focus on the daily scheduling of therapy services, henceforth denoted as therapy jobs, hospital-wide. In hospitals, every morning of a planning day, therapy jobs have to be assigned to rooms, therapists and start times. Unlike in rehabilitation hospitals, it has to be decided if a therapy job will be executed at a hospital's therapy center (TC), which consists of differently equipped rooms, or at a ward, where inpatients stay. Limited resources such as TC rooms and therapists have to be considered, and labor requirements such as breaks are to be taken into account. The hospital strives to minimize lateness with respect to the earliest possible start time of therapy jobs. We model the problem using Integer Programming (IP) and compare its solutions with those provided by a cutting plane and a heuristic algorithm.

Within the health care planning matrix of Hans et al. (2011), the problem can be classified as offline operational resource capacity planning. Our approaches thus rely on planning decisions which have been made at higher planning levels, such as strategic decisions on the number of TC rooms or tactical decisions e.g. on the shift schedule of therapists. Therefore, we assume given shift-assignments for each day under consideration. As the therapy jobs become known at the beginning of each planning day, the scheduler assigns the available therapists with their predefined shifts on the specific day to patients, rooms and times to ensure that all treatments are undertaken. Moreover, in contrast to the planning of home care services, where therapists or nurses visit patients at their home, some inpatients walk or can be transported to the TC. Another difference to home care is that patients require a much broader variety of therapist skill levels or have specific therapy room requirements.

In order to solve the problem, we present an Integer Program (IP) which assigns therapists to jobs within the jobs' allowed time windows. It considers precedence relations between different jobs (for one patient), and handles different job durations and job requirements such as required qualifications. Besides the job constraints, the model takes into account the walking distances of therapists within the hospital, and their individual shift regulations with given shift start time, shift duration as well as at least one break, as regulated by law. The scheduling character of the problem motivated us to model the problem as a variant of the multi-mode resource-constrained project scheduling problem (MMRCPSP) which is known to be NP-hard (Kolisch and Drexl (1997)). In practice, patient demand for a given day is known during the evening prior to it, and jobs for the given day must all be scheduled by 6 a.m. As a consequence, computation time has to be less than 10 hours. Our computational study will reveal that computation times for solving practice-relevant problems may become very long. As a consequence, we developed a solution procedure which breaks down the problem into two parts: First, we assign each job to a therapist, room and start time. The solution of this problem, which we denote as therapist scheduling problem, is then checked in the second part with respect to feasibility of therapists' travel times. If the schedule turns out to be infeasible, a constraint (i.e., a cutting plane) is added to the therapist scheduling problem. The resulting cutting plane algorithm provides exact solutions for practice-relevant problems. However, it does not guarantee short solution times. To account for the need for short solution times, we develop a sequential allocation heuristic. Experimental results reveal that hospitals can use the cutting plane algorithm to solve real-world hospital-wide therapist scheduling and routing problems to optimality within an acceptable computation time. For larger problem sizes, the heuristic approach generates good quality solutions in a fraction of the time required by the mathematical program and the cutting plane algorithm.

The remainder of this paper is structured as follows. Section 2 provides an overview of related work. Section 3 presents the mathematical model, followed by the cutting plane algorithm in Section 4. In Section 5, the sequential allocation approach is presented. The results of the experimental study are shown in Section 6. Section 7 provides a discussion of the limitations of our approaches followed by a conclusion.

2 Related work

There is a rich literature on both, scheduling and routing problems in healthcare. Operating theater, outpatient appointment, radiotherapy, physician, patient and integrated *scheduling* problems have been reviewed by and in Cardoen et al. (2010), Ahmadi-Javid et al. (2017), Vieira et al. (2016), Erhard et al. (2018), Gartner and Padman (2017) and Marynissen and Demeulemeester (2016), respectively. *Routing* of ambulances and patients have been reviewed in Hulshof et al. (2012). Literature on *routing and scheduling* with a focus on home health care is reviewed by Fikar and Hirsch (2017) and Cissé et al. (2017). Routing of home care nurses is highly relevant for our paper because nurses typically visit patients periodically, in order to perform therapeutic treatments. This problem is similar to the structure of the problem considered in our paper where physical therapists visit intensive care patients who cannot be transported. In the following, we will extend our focus towards physical therapy planning as well as routing problems in hospitals. The most relevant papers for this research will be reviewed and a statement to what extent they differentiate from our work will be given.

A considerable number of papers focus on physical therapy planning in hospitals. Griffiths et al.

(2012), for example, formulate a physical therapy scheduling problem as a multi-objective combinatorial optimization problem. Motivated by an inefficient manual approach to create schedules, they solved the problem using a three stage local search solution procedure. The novelty of our approach as compared to their problem is that we take into account routing decisions and develop not only a heuristic but also an optimal solution approach.

Another therapy scheduling paper is the work of Schimmelpfeng et al. (2012). They combine the strategic, tactical and operational planning level and use a three-stage Mixed-Integer Program (MIP) to maximize the number of treatments. Our work is different because we have to ensure that all treatments are scheduled while distances in the hospital between the treatment locations require routing of therapists.

A smaller number of papers focus on routing problems in hospitals. Beaudry et al. (2010) consider a dynamic transportation problem of patients in a hospital which is based on the well-known dial-a-ride problem. The major difference of our problem is that we route physical therapists within a hospital by taking decisions on the location of physical therapy treatments. In Beaudry et al. (2010) these locations are fixed.

Similarly, Hanne et al. (2009) focus on dynamic transportation of patients with different heuristic strategies to assign transportation jobs to vehicles. In contrast, in this paper, we focus on jobs which are known in advance and have to be scheduled at the beginning of each planning day.

Schmid and Doerner (2014) consider the problem of scheduling examinations, surgeries and routing patients, simultaneously. They use a metaheuristic approach to solve the problem. While one difference with respect to this problem is that we route physical therapists instead of patients, we additionally present an exact solution procedure.

Doulabi et al. (2016) focus on operating room planning and scheduling. Similar to our approach, the maximum daily working hours of staff is considered and the same member of staff cannot be assigned to two activities at the same time. Also, they provide an infeasibility detection algorithm. The difference with our work is, however, that their algorithm checks infeasibilities based on the multidimensional knapsack problem whereas we check individual therapists' routes.

Roshanaei et al. (2017a) and Roshanaei et al. (2017b) focus on multi-hospital surgical case scheduling for operating rooms. The authors model the problem using Integer Programming and use a logic-based Benders' method to solve it. Their problem differs from our problem because they assume that clinicians have been assigned to operating rooms based on the Master Surgical Schedule. Accordingly, routing of staff is irrelevant in their problem. However, from a methodological point of view, the major similarity to our paper is that they incorporate feasibility and optimality cuts into their problem in order to speed up the solution process.

Apart from the healthcare-related *scheduling* work, the resource-constrained project scheduling problem is one of the most relevant for our work. Koné et al. (2011) provide different mathematical programming formulations for this problem, which is related to the model we provide in Section 3. More specifically, the multi-mode version of the problem is most related and has been formulated by Talbot (1982) using binary programming and more recently, lifting constraints have been applied successfully to speed up the solution process as demonstrated in Delorme (2017). The major difference to the classical single- and multi-mode resource-constrainted project scheduling work is, however, that we have a combination between scheduling and routing which we formulate as a mathematical program.

Apart from the healthcare-related *routing* work, the capacitated vehicle routing problem with time windows is one of the most relevant for us. A MIP formulation is provided by Kang et al. (2008). The major difference to our work is, however, that we have a combination between scheduling and routing decisions where the location to carry out the therapy job is, in some cases, flexible. Providing a mathematical model and solution procedures for this problem type is a major difference to previous work in vehicle routing problems.

Summarizing, the model and solution approaches proposed in this paper can be categorized into and differentiated from the literature as follows. An aggregated approach is followed in which patients' therapy jobs or surgeries are assigned to start times. This approach is similar to Schimmelpfeng et al. (2012), Griffiths et al. (2012), Roshanaei et al. (2017a) and Roshanaei et al. (2017b). However, these approaches do not take into account routing decisions for staff. Secondly, with respect to the objective function, our approach has a patient-oriented objective function where treatments are scheduled as early as possible during the day. Thirdly, with respect to the constraints, the model combines the approaches of Schimmelpfeng et al. (2012) and Griffiths et al. (2012) by considering a number of real-world constraints such as precedence constraints, time windows, breaks for therapists and preferences of patients. Finally, with respect to the solution approach, Roshanaei et al. (2017a) and Roshanaei et al. (2017b) are most relevant because of the cutting planes applied to the mathematical program while in our problem, the application area is different and we have both a routing and a scheduling problem.

3 Problem description and model formulation

In this section, we first provide a problem description followed by a formal introduction of the sets, indices and the model formulation. The therapist scheduling and routing problem (ThSRP) is interpreted as a MMRCPSP (Kolisch and Drexl (1997)) while additional constraints are formulated with respect to the therapists' working regulations and routing requirements.

Every morning of a planning day, therapy jobs have to be assigned to rooms, therapists and start times. A scheduler has to decide if a job will be executed at a hospital's therapy center, which consists of differently equipped rooms, or at a ward, where inpatients recover from surgery. Sometimes, there is no flexibility in the job-to-room assignment, for example, when intensive care unit patients receive a physical therapy. Also, it can happen that there is only one therapist skilled to carry out a particular job. All the limited resources such as therapy rooms and therapists' skills have to be considered by the scheduler. In addition, labor requirements such as breaks have to be taken into account. Now, when jobs are assigned to rooms, therapists may need to travel between different locations. The scheduler strives to achieve the hospital's objective which is to minimize lateness with respect to the earliest possible start time of therapy jobs.

3.1 A formal problem description

Planning horizon and therapists Let $\mathcal{T} = \{0, \ldots, T\}$ be the set of discrete times, within the planning horizon T of one day, which are equally incremented (typically, 15 minutes). In the following, we will relate to the time span between two adjacent times as a period. Set $\mathcal{P}$ depicts a set of therapists where the shift within a working day of therapist $p \in \mathcal{P}$ starts at time S_p^{sh}. The working day of a therapist is between the regular working time of Δ^{regW} periods and the maximal allowed working time of Δ^{maxW} periods. As required by law, during the regular working time, each therapist has to have one break. When the regular working time of a therapist is exceeded, which happens when there exists a time $S_p^{\text{sh}} + \Delta^{\text{regW}} \leq t \leq S_p^{\text{sh}} + \Delta^{\text{maxW}}$ in which a job is scheduled for therapist $p \in \mathcal{P}$, the therapist has to have a second break. Therapist p's possible times for the shift end are given by set $\mathcal{S}_p^{\text{e}}$. Naturally, therapists have different qualifications required for different treatments of patients.

Rooms and travel times between rooms Therapists are routed within the hospital and visit a set of rooms $\mathcal{R} = \mathcal{R}^{\text{icu}} \cup \mathcal{R}^{\text{tc}} \cup \mathcal{R}^{\text{ward}} \cup \{0\}$ which can be broken down by ICU ($\mathcal{R}^{\text{icu}}$), therapy center TC ($\mathcal{R}^{\text{tc}}$) and ward rooms ($\mathcal{R}^{\text{ward}}$) . Therapists take their break in break room $r = 0$. TC rooms are specially equipped and relevant for outpatients. In addition, inpatients who recover on a ward can walk to a TC room where they receive their treatment. For ICU patients, the therapy takes place within the ICU. The time required by one therapist to travel from room $r \in \mathcal{R}$ to room $s \in \mathcal{R}$ is $c_{r,s}$.

Therapy jobs Therapy jobs, henceforth called jobs, represent treatments, and patients can have multiple treatments during a day. Therapy job set $\mathcal{I} = \mathcal{I}^{\text{icu}} \cup \mathcal{I}^{\text{tc}} \cup \mathcal{I}^{\text{ward}}$ comprises ICU, TC and ward jobs. An ICU job $i \in \mathcal{I}^{\text{icu}}$ has to be executed in its designated ICU room $r \in \mathcal{R}^{\text{icu}}$, a ward job $i \in \mathcal{I}^{\text{ward}}$ can be executed either in the corresponding ward room $r \in \mathcal{R}^{\text{ward}}$, or at a room $r \in \mathcal{R}^{\text{tc}}$, within the hospital's TC. An outpatient job $i \in \mathcal{I}^{\text{tc}}$ has to be executed in one TC room $r \in \mathcal{R}^{\text{tc}}$.

Each job $i \in \mathcal{I}$ has a unique therapy duration d_i^{job} which is independent of the therapist and the room where the therapy takes place. Each job can be performed in different modes (for the modelling concepts of modes see, e.g., Kolisch and Drexl (1997)) where a mode specifies the therapist providing the treatment and the required room. Set $\mathcal{M}_i$ gives the set of all modes in which job $i \in \mathcal{I}$ can be executed. Binary parameters $r_{i,m,r}^{\text{room}}$ and $r_{i,m,p}^{\text{th}}$ are 1 if job i in mode m requires room k and therapist p, respectively. Using this concept we enumerate for each job all possible combinations of therapists and rooms. In Section 3.5 we provide an example with 2 rooms and 1 therapist.

Each job i can be started at any time within the set $\mathcal{W}_i$. For example, $\mathcal{W}_1 = \{1, 2, 3, 12, 20\}$ denotes that job $i = 1$ can be started, in time 3 but not in time 11. By E_i^{job} and L_i^{job} we indicate the earliest and latest start time of job i. The end of the planning horizon is determined by the maximum latest job finish time plus the maximum travel time between two rooms $T = \max\limits_{i \in \mathcal{I}} \left\{ L_i^{\text{job}} + d_i^{\text{job}} \right\} + \max\limits_{r,s \in \mathcal{R}} \{c_{r,s}\}$.

Break jobs We introduce break jobs $i_{1,p}^{\text{br}}$ and $i_{2,p}^{\text{br}}$, which represent the first and second break of therapist p, respectively. To ensure that breaks are scheduled in the middle of a shift, we assign time windows to these jobs. The time windows prohibit that breaks are scheduled in the beginning or at the end of a shift.

Resource capacities Scarce resources are rooms and therapists. Each room $r \in \mathcal{R}$ has a capacity of K_r which relates to the number of therapy jobs which can take place in the room. Ward rooms have a capacity of 1, while rooms in the therapy center with specialized workout equipment have a capacity greater than 1. Each therapist $p \in \mathcal{P}$ has a capacity of 1 during his working time if he is not taking a break. In contrast to the rooms where the capacity is constant and given, the availability of the therapists is decided upon in the mathematical program.

Therapy pathways Sometimes, a patient requires multiple therapies within a day. Since we know the patient behind each job, we can model a therapy pathway. Therapy pathways, similar to clinical pathways (van de Klundert et al. (2010)), are typically evidence-based health care processes. We introduce these pathways in order to avoid patient harm if consecutive treatments follow each other too quickly. A therapy pathway is a sequence of therapy jobs with minimum time lags $d_{i,j}^{\min} \in \mathbb{Z}_{\geq} 0$ between the start of job i and its successor j. Let $\mathcal{E} \subseteq \mathcal{I} \times \mathcal{I}$ denote the set of all precedence relations between jobs. A minimum time lag $d_{i,j}^{\min}$ stipulates that job j has to be started at least $d_{i,j}^{\min}$ periods later than the start of job i.

Objective The objective of our problem is to schedule jobs in the earliest possible time. The rationale for choosing this objective is three-fold: First, if a treatment starts as early as possible, recovery can start early. Second, treating a patient as early as possible reduces patient waiting times for the therapy. Finally, from a therapist's perspective, it is desirable to treat patients earlier rather than later during the day to ensure an early finish of the therapists' working day.

3.2 Introduction of a multi-mode concept for therapist and room requirements

We employ multiple modes to reflect qualification and room requirements for each job as follows. Let $\mathcal{M}_i$ denote the set of modes in which job $i \in \mathcal{I}$ can be executed. For job $i \in \mathcal{I}$ and mode $m \in \mathcal{M}_i$ we denote the capacity requirement for room $r \in \mathcal{R}$ with $r_{i,m,r}^{\text{room}}$ and for therapist $p \in \mathcal{P}_i$ with $r_{i,m,p}^{\text{th}}$. Binary parameter $r_{i,m,r}^{\text{room}} = 1$ if job $i \in \mathcal{I}$ requires room $r \in \mathcal{R}$ in mode $m \in \mathcal{M}_i$ and 0 otherwise. Similarly, $r_{i,m,p}^{\text{th}} = 1$ if job $i \in \mathcal{I}$ requires therapist $p \in \mathcal{P}$ in mode $m \in \mathcal{M}_i$ and 0 otherwise. Using this concept we can enumerate all the possible combinations between therapists and rooms. An advantage of this concept is that it can be employed to take into account therapist qualifications. For example, if a job requires a therapist with a specific qualification, then $r_{i,m,p}^{\text{th}} = 1$ only for the therapists that possess this qualification.

3.3 Model formulation

Having introduced the parameters and the mode concept, we now turn to the introduction of the decision variables and their domains as well as the constraints.

3.3.1 Decision variables and objective

We employ the following four binary decision variables:

$$x_{i,m,t} = \begin{cases} 1, \text{ if job } i \in \mathcal{I} \text{ starts in mode } m \in \mathcal{M}_i \text{ at time } t \in \mathcal{W}_i \\ 0, \text{ otherwise} \end{cases} \tag{1}$$

$$y_{p,t} = \begin{cases} 1, \text{ if therapist } p \text{ is working at time } t \in \{S_p^{\text{sh}}, S_p^{\text{sh}}+1, \ldots, S_p^{\text{sh}} + \Delta^{\text{maxW}}\} \\ 0, \text{ otherwise} \end{cases} \tag{2}$$

$$z_{i,j} = \begin{cases} 1, \text{ if job } i \in \mathcal{I} \text{ is scheduled before job } j \in \mathcal{I} \\ 0, \text{ otherwise} \end{cases} \tag{3}$$

$$w_{i,j,p,r,s} = \begin{cases} 1, \text{ if job } i \text{ and job } j \text{ are assigned to therapist } p \text{ with job } i \text{ assigned to room } r \\ \quad \text{and job } j \text{ assigned to room } s \\ 0, \text{ otherwise} \end{cases} \tag{4}$$

The objective function (5) minimizes the sum of the activity start times and thus aims at treating patients as early as possible within the treatment time windows. This is equivalent to minimizing total waiting time of patients:

$$\text{Minimize } z = \sum_{i \in \mathcal{I}} \sum_{m \in \mathcal{M}_i} \sum_{t \in \mathcal{W}_i} t \cdot x_{i,m,t} \tag{5}$$

3.3.2 Constraints

In the following, we model the constraints for the physical therapy scheduling and therapist routing problem.

Mode and time assignment, time lag and capacity constraints To ensure that jobs are performed and therapy pathways and resource capacities are respected, the following sets of constraints (6)–(9) are required:

$$\sum_{m \in \mathcal{M}_i} \sum_{t \in \mathcal{W}_i} x_{i,m,t} = 1 \qquad \forall i \in \mathcal{I} \tag{6}$$

$$\sum_{m \in \mathcal{M}_i} \sum_{t \in \mathcal{W}_i} t \cdot x_{i,m,t} + d^{\min}_{i,j} \leq \sum_{m \in \mathcal{M}_j} \sum_{t \in \mathcal{W}_j} t \cdot x_{j,m,t} \qquad \forall (i,j) \in \mathcal{E} \tag{7}$$

$$\sum_{i \in \mathcal{I}} \sum_{m \in \mathcal{M}_i} r^{\text{room}}_{i,m,r} \sum_{\tau=\max\{E^{\text{job}}_i, t-d^{\text{job}}_i+1\}}^{\min\{L^{\text{job}}_i, t\}} x_{i,m,\tau} \leq K^{\text{room}}_r \qquad \forall r \in \mathcal{R}, t \in \mathcal{T} \tag{8}$$

$$\sum_{i \in \mathcal{I}} \sum_{m \in \mathcal{M}_i} r^{\text{th}}_{i,m,p} \sum_{\tau=\max\{E^{\text{job}}_i, t-d^{\text{job}}_i+1\}}^{\min\{L^{\text{job}}_i, t\}} x_{i,m,\tau} = y_{p,t} \qquad \forall p \in \mathcal{P}, t \in \{S^{\text{sh}}_p, S^{\text{sh}}_p + 1, \ldots, S^{\text{sh}}_p + \Delta^{\text{maxW}}\} \tag{9}$$

Constraints (6) ensure that each job i is assigned to one mode and one time within its time window. Constraints (7) ensure minimum time lags between jobs. Constraints (8) are the capacity constraints for rooms. Constraints (9) set the y-variables (therapist p is working at time t) to 1 whenever a job j is scheduled to be processed at t and requires therapist p. In the previous two constraints, the index set of the third sum is defined as $\tau = \max\{E^{\text{job}}_i, t - d^{\text{job}}_i + 1\} \ldots \min\{L^{\text{job}}_i, t\}$. The bounds can be explained using the following three examples: First, let period t be earlier than the earliest start of the therapy job E_i. In this case, the constraint does not bind as the index set is empty because of the lower bound being larger than the upper bound. Second, if period t is later than the latest finish of therapy job L_i, then the constraint does not bind, either. The term $t - d^{\text{job}}_i + 1$ would be later than L_i. Third, periods that are between the jobs' start time and the latest finish time are included in the index set which ensures that the constraint is binding.

Working time regulations Working time regulations are presented in constraints (10)–(12):

$$\sum_{m \in \mathcal{M}_{i^{\text{br}}_{1,p}}} \sum_{t \in \mathcal{W}_{i^{\text{br}}_{1,p}}} x_{i^{\text{br}}_{1,p},m,t} = 1 \qquad \forall p \in \mathcal{P} \tag{10}$$

$$\sum_{m \in \mathcal{M}_{i^{\text{br}}_{2,p}}} \sum_{t \in \mathcal{W}_{i^{\text{br}}_{2,p}}} x_{i^{\text{br}}_{2,p},m,t} \leq 1 \qquad \forall p \in \mathcal{P} \tag{11}$$

$$\sum_{t=S^{\text{sh}}_p+\Delta^{\text{regW}}}^{S^{\text{sh}}_p+\Delta^{\text{maxW}}} y_{p,t} \leq M \sum_{m \in \mathcal{M}_{i^{\text{br}}_{2,p}}} \sum_{t \in \mathcal{W}_{i^{\text{br}}_{2,p}}} x_{i^{\text{br}}_{2,p},m,t} \qquad \forall p \in \mathcal{P} \tag{12}$$

Constraints (10) ensure that the first break of a therapist is scheduled exactly once. For example, assume we have one therapist with an 8 hour working day and shift start $S^{\text{sh}}_1 = 1$. Setting the break job time window to $\mathcal{W}_{i^{\text{br}}_{1,1}} = \{17\}$ ensures that the first break starts exactly in period 17 which is after the completion of 16 15-minutes time increments or 4 hours of work. More precisely, Constraints (10) ensure that $x_{i^{\text{br}}_{1,1},1,17} = 1$.

Constraints (11) guarantee that no more than one second break is scheduled. Constraints (12) ensure that if a therapist p works beyond his regular working time, he must have a second break. For these constraints, we use big $M = \Delta^{\text{maxW}} - \Delta^{\text{regW}}$.

Therapist routing constraints The following constraints ensure that the traveling time of a therapist is taken into account when planning consecutive jobs in different rooms:

$$\sum_{\substack{m \in \mathcal{M}_i: \\ r^{\text{th}}_{i,m,p} = r^{\text{room}}_{i,m,r} = 1}} \sum_{t \in \mathcal{W}_i} x_{i,m,t} + \sum_{\substack{m \in \mathcal{M}_j: \\ r^{\text{th}}_{j,m,p} = r^{\text{room}}_{j,m,s} = 1}} \sum_{t \in \mathcal{W}_j} x_{j,m,t} \leq 1 + w_{i,j,p,r,s} \quad \forall i \in \mathcal{I} \cup \mathcal{I}^{br}, j \in \mathcal{I} \cup \mathcal{I}^{br}, p \in \mathcal{P}, r \in \mathcal{R}, s \in \mathcal{R}, i \neq j, r \neq s \tag{13}$$

$$\sum_{m \in \mathcal{M}_i} \sum_{t \in \mathcal{W}_i} t \cdot x_{i,m,t} - \sum_{m \in \mathcal{M}_j} \sum_{t \in \mathcal{W}_j} t \cdot x_{j,m,t} \geq -z_{i,j} \cdot M \quad \forall i \in \mathcal{I} \cup \mathcal{I}^{br}, j \in \mathcal{I} \cup \mathcal{I}^{br}, i \neq j \tag{14}$$

$$\sum_{m \in \mathcal{M}_j} \sum_{t \in \mathcal{W}_j} t \cdot x_{j,m,t} - \left(\sum_{m \in \mathcal{M}_i} \sum_{t \in \mathcal{W}_i} t \cdot x_{i,m,t} + d^{\text{job}}_i + c_{r,s} \right) \geq (z_{i,j} \cdot w_{i,j,p,r,s} - 1) \cdot M \quad \forall i \in \mathcal{I} \cup \mathcal{I}^{br}, j \in \mathcal{I} \cup \mathcal{I}^{br}, p \in \mathcal{P}, r \in \mathcal{R}, s \in \mathcal{R}, i \neq j, r \neq s \tag{15}$$

Constraints (13) set the w-variable to 1 if therapist p is assigned to job i taking place in room r, and job j taking place in room s. Next, constraints (14) set the z-variable to 1 if job i is scheduled before job j. Finally, constraints (15) consider the travel time for therapist p who is assigned to job i in room r, and job j in room s with job i, performed before job j. We set big $M = \max_{i \in \mathcal{I}} L^{\text{job}}_i - \min_{i \in \mathcal{I}} E^{\text{job}}_i$. Since constraints (15) are non-linear, we linearized them using the procedure described in Section B in the appendix.

3.4 A lower bound for the Therapist Scheduling and Routing Problem

By relaxing the routing constraints (13)–(15) of the therapist scheduling and routing problem (1)–(15) we obtain the therapist scheduling problem (ThSP). The ThSP reads as follows:

$$\text{Minimize } z$$
$$\textit{subject to Constraints } (6) - (12)$$
$$\textit{in which } z \textit{ is defined in Equation } (5)$$
$$\textit{and decision variables are defined in Equations } (1) - (4).$$

Obviously, solving the ThSP to optimality provides a lower bound of the optimal solution of the ThSRP.

3.5 An example

In what follows we provide a short example to illustrate the model. We first give a description of the parameters, followed by the number of decision variables and constraints. Finally, we provide the optimal solution of the problem.

3.5.1 Parameters

We have a set of times $\mathcal{T} = \{0, 1, \ldots, 15\}$ where each time $t \in \mathcal{T}$ represents the start of the period of 15 minutes length following t. We consider a single therapist $p = 1$ who is qualified for all jobs. The hospital has a break room, a therapy center consisting of one room, one ward room and one ICU room indexed by $r = 0, 1, 2, 3$. The set of rooms thus is $\mathcal{R} = \{0, 1, 2, 3\}$. Assume travel durations $c_{r,s} = 1$ between rooms $r, s \in \mathcal{R} : r \neq s$. The set of jobs related to physical therapies is $\mathcal{I} = \{1, 2, 3, 4\}$ with sets of times $\mathcal{W}_1 = \{0, 1, \ldots, 6\}$, $\mathcal{W}_2 = \{9, 10, \ldots, 14\}$, $\mathcal{W}_3 = \{0, 1, \ldots, 5\}$ and $\mathcal{W}_4 = \{9, 10, \ldots, 13\}$. Durations of jobs are $d_1^{\text{job}} = d_2^{\text{job}} = 2$ and $d_3^{\text{job}} = d_4^{\text{job}} = 3$. The therapist has one fixed break and one flexible break, both of them have to be performed in the break room. Accordingly, the first and the second break are represented by jobs 5 and 6. Both break jobs have a duration of $d_5^{\text{job}} = d_6^{\text{job}} = 1$. Room capacities are $R_{k,t} = 1 \quad \forall k \in \mathcal{R}, t \in \mathcal{T}$. Time lags are $d_{1,2}^{\min} = d_{3,4}^{\min} = 8$ and mode-dependent room resource requirements are $r_{1,1,1}^{\text{room}} = 1, r_{1,2,2}^{\text{room}} = 1, r_{2,1,1}^{\text{room}} = 1$, $r_{3,1,1}^{\text{room}} = 1$ and $r_{4,1,3}^{\text{room}} = 1$. This reveals that job $i = 1$ has $|\mathcal{M}_1| = 2$ modes such that the job can be executed either in the TC (room $r = 1$) or on the ward (room $r = 2$). All other jobs have fixed locations. Since we have only one therapist, the resource requirements come up to $r_{i,1,1}^{\text{th}} = 1 \quad \forall i \in \mathcal{I}$.

3.5.2 Solution

The number of decision variables and constraints for this example is 4,056 and 2,920, respectively, and CPLEX solves this problem in 0.05 seconds. Figure 1 provides a graphical representation of the schedule using a Gantt chart.

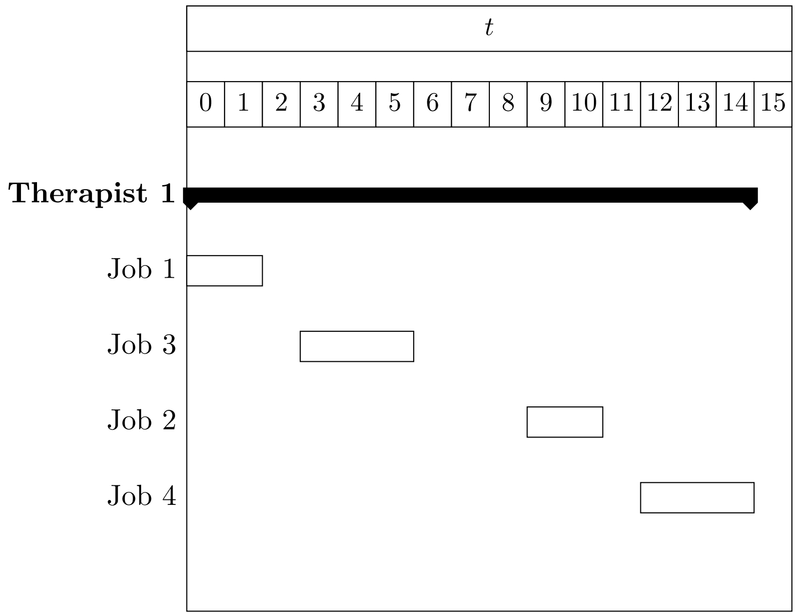

Figure 1: Therapist 1's schedule for the example

4 A cutting plane algorithm

Cutting plane algorithms have been applied widely in Integer Programming (Nemhauser and Wolsey (1999)). To make problem (1)–(15) more tractable when handling real-world instances, we iteratively solve it using cutting planes. Section 4.1 and 4.2 describe the general framework and the scheme to generate cuts. As we will see the procedure will lead to the optimal solution of the ThSP.

4.1 Cutting plane framework

Figure 2 presents the general framework in which the ThSP introduced in Section 3.4 is incorporated into the initial stage. The set of cutting planes (17) which are detailed in Section 4.2 is initially empty and new constraints are added iteratively throughout the solution procedure. This procedure is similar to adding feasibility or optimality cuts in Stochastic Integer Programs (Laporte and Louveaux (1993)).

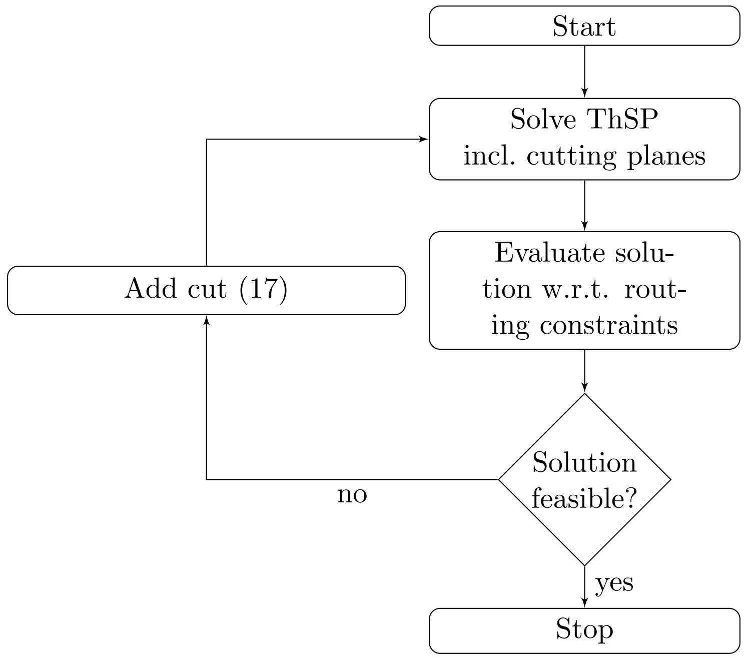

Figure 2: Decomposition scheme

At each iteration,the ThSP is solved and if its solution turns out to be infeasible with respect to the routing constraints (see Section 4.2), we insert additional cutting planes and the procedure starts over. We terminate this iterative procedure until the first master problem turns out to be routing-feasible.

4.2 Generation of cutting planes

Based on the solution of the ThSP we introduce cuts by first calculating job-room-start time tuples followed by checking infeasibility and constraint generation.

Having obtained an optimal solution $\mathcal{X}^* := \{x^*_{i,m,t} | x^*_{i,m,t} = 1 \quad \forall i, m, t\}$ of the ThSP we can generate for each therapist $p \in \mathcal{P}$ the set $\mathcal{L}_p$. Each tuple $(i, r, t) \in \mathcal{L}_p \subseteq \mathcal{I} \times \mathcal{R} \times \mathcal{T}$ consists of a job which has to be performed by therapist p at time t in room r. Algorithm 1 in Appendix C.1 gives a formal description for generating $\mathcal{L}_p$ from $\mathcal{X}^*$. We now order therapists' jobs according to increasing start times and check for each pair of succeeding jobs (i, r, t) and (i', r', t') with $i \neq i'$ and $t' > t$ whether the routing is feasible, i.e., if the following condition holds true:

$$t + d_i^{\text{job}} + c_{r,r'} \leq t' \quad \forall p \in \mathcal{P}, (i, r, t), (i', r', t') \in \mathcal{L}_p \tag{16}$$

Once the solution $\mathcal{X}^*$ is detected infeasible, i.e., Conditions (16) are not satisfied, we blacklist the corresponding solution by adding cut

$$\sum_{x^*_{i,m,t} \in \mathcal{X}^*} x^*_{i,m,t} \leq |\mathcal{X}^*| - 1 \tag{17}$$

into the ThSP. The cardinality constraint (17) represents a cutting plane which allows to set at most $(|\mathcal{X}^*| - 1)$ x-variables of the ThSP's solution previously assigned to be 1 and, thus, excludes solution $\mathcal{X}^*$ from the polyhedron (Stephan (2010)). This procedure is similar to subtour elimination constraints in the Travelling Salesman Problem.

The objective function of the the relaxed version of the ThSP with no or only a subset of cuts of the form (17) truly yields an upper bound for the original ThSP with routing constraints. Adding one cut just forbids the current solution yielding to an infeasible solution for the routing constraints. By adding as many of such cuts as necessary to obtain a feasible solution for the ThSP we can state:

Proposition 1 *The cutting plane algorithm with cuts (17) leads to the optimal solution of the ThSP.*

Proof. Recall from inequality (16), we have a finite set of therapists $\mathcal{P}$ and a finite set of consecutive job-room-start-time combinations for each therapist $p \in \mathcal{P}, (i, r, t), (i', r', t') \in \mathcal{L}_p$. Accordingly, the number of cuts which can be added is finite. As a consequence, in each iteration a new cut which was not added before is added to the set of cuts. Therefore, after a finite number of iterations, the algorithm will find an optimal solution. □

5 A sequential allocation heuristic (SAH)

Since large-sized problems are intractable due to the complexity of the problem, we propose a sequential allocation heuristic (SAH). The SAH consists of two stages and takes advantage of the fact that a subset of patients (outpatients) have scheduled appointments several days or weeks before the actual treatment. For other patients (inpatients), however, job start times are flexible. We decompose the schedule generation into an assignment of jobs with fixed start times followed by the assignment of jobs with flexible start times given in Algorithms 2 and 3 in the Appendix, respectively.

5.1 Stage 1: Assign jobs with fixed start times

In the outer loop of Algorithm 2 (see Section C.2 in the Appendix), therapy jobs that have a fixed start are assigned to therapists and rooms. In the inner loop (Lines 3–20), the algorithm runs over the planning horizon $\mathcal{T}$. Once the earliest start time is reached for a fixed therapy job i, i.e., $t = E_i^{\text{job}}$, the algorithm iterates over all therapists $p \in \mathcal{P}$ which are ordered by their shift starts (Lines 5–14). In Line 6 it is ensured that the therapist p under consideration meets job i's required qualifications. The mode matrix ensures that the qualification is met. More formally, the condition $m \in \mathcal{M}_i : r_{i,m,p} = 1$ has to hold. The next inner loop (Lines 7–13) looks for a suitable room r for job i in increasing order of room index. Finally, lines 9, 10 and 11 are the assignment step and the update of the available resource capacity, respectively.

5.2 Stage 2: Assign jobs with flexible start times

Algorithm 3 (see Section C.3 in the Appendix), describes the second part of the sequential allocation heuristic in which therapy jobs with flexible start times, i.e., $i \in \mathcal{I} : E_i^{\text{job}} < L_i^{\text{job}}$ are scheduled. The algorithm is similar to scheduling therapy jobs with fixed time. However, the major difference is Line 3 in which the set of start times $\mathcal{W}_i$ of therapy job i is considered.

Similar to the mathematical program introduced in Section 3, breaks are taken into account by using break jobs. Also a mode matrix, see Section 3.2, is used which ensures that break jobs are executed by therapists in the break room.

6 Experimental study

The experimental study is organized as follows: In Section 6.1 we explain how we obtained the hospital data and how we set up the parameters followed by an introduction of the test instances in Section 6.2. A description of the evaluation measures and the computation time analysis are given in Section 6.3. In this section, we show how we tested our Integer Program, the cutting plane algorithm and the SAH and give a comparison of these methods. We provide an evaluation of lateness and the number of unscheduled jobs in Section 6.4.

6.1 Hospital data

We collected data from a 350-bed hospital in Germany. Between May 22nd and June 28th 2013, a total of 27 schedules were collected. Weekends and holidays were excluded. On the supply side of the data, the therapists' working pattern, their skills, room information including equipment, among others, were registered. On the demand-side, for each therapy job its skill, room requirements and time windows were collected. Across the 27 days in which data was collected, the number of inpatient jobs that had to be planned varied between 91 and 136. Also, data about therapy pathways between jobs were collected. The hospital's layout was analyzed separately in order to collect data on walking distances. Based on the legal working time regulations, we set up parameters such as the maximum working time allowed between the start of the therapists' working day and the first and second break.

6.2 Instance generation

For the generation of the test instances, instances 6, 7 and 8 were randomly selected from the above mentioned data collection period. Instance 6 represents a less busy day while instances 7 and 8 represent data from two busy days. The major difference is, however, that the number of therapists was on one day lower than on the other. Instance 5 was generated by reducing the size of instance 8 by $1/3$. Instances 1–4 were generated randomly, however, following the hospital's layout, skill mix and working patterns, among others. We set the planning horizon of length $T = 16$ and $|\mathcal{R}| = 5$ rooms including the break room. Naturally, the collaborating hospital has more than 5 rooms. However, they can be aggregated as follows:

Break room In this room, breaks take place and no aggregation had to be performed.

ICU The unit consists of multiple bays which were aggregated into one ICU room.

TC rooms All six TC rooms, the therapy pool, training hall and backup therapy room were aggregated as one room.

Ward rooms The ward rooms were aggregated into two types of rooms: One where multiple treatments can take place and a second one where the capacity to carry out therapies is limited to 1.

Table 1 provides an overview of the test instances that we generated.

Instance	# Jobs	# Therapists	# Modes
1	16	3	81
2	30	5	260
3	46	8	616
4	58	9	918
5	80	8	1,296
6	104	12	2,424
7	120	10	2,520
8	120	12	2,904

Table 1: Test instances

The table shows that comparing test instances 1 and 5, a 5 times increase of jobs and a 2.7 times increase of therapists results in an increase of modes by factor 16. Comparing instance 3 and 5, in which the number of therapists is identical, a 1.7 times increase of jobs more than doubles the total number of modes.

6.3 Computational results

The mathematical program, the cutting plane and the heuristic algorithm are evaluated in terms of solution time and objective function value. Since the SAH does not guarantee that all jobs are scheduled, we also report the number of unscheduled jobs. For the IP and the cutting plane algorithm, we additionally report the number of decision variables and constraints. More specifically, we report the initial number of decision variables and constraints in the beginning and in the end of the cutting plane algorithm. Finally, we report its number of iterations.

All computations were performed on a 2.4 GHz Intel Core i7 (4700MQ) and 32 GB RAM of which we allocated 25 GB to the Java Virtual Machine. The models were coded in Java using the 64 bit version of the application programming interface of ILOG CPLEX 12.6. For computational comparison, we coded the heuristic algorithm in Java, too. Table 2 provides an overview of the computational results with the performance metrics of the full formulation of the Integer Program (IP) and the cutting plane algorithm (CP). An asterisk (*) means that although CPLEX was able to build the model, it failed to start preprocessing and thus failed to find a solution. # Iterations refers to the number of iterations for the cutting plane algorithm. It is equal to the number of constraints added to the therapist scheduling and routing problem (ThSRP).

	# Variables		# Constraints			Solution time [ms]		
Instance	IP	CP	IP	CP	# Iterations	IP	CP	SAH
1	528	528	137	135	2	50	54	1
2	14,356	1,360	4,227	186	77	362,959	2,564	1
3	38,185	3,376	7,578	257	0	4,535,760	83	1
4	54,107	5,526	15,569	288	194	9,499,669	37,909	1
5	102,178*	7,600	27,768*	294	9	–*	9,396	1
6	189,450*	14,736	45,122*	392	153	–*	170,625	1
7	145,360*	8,020	39,494*	374	276	–*	66,648	1
8	254,471*	16,080	71,172*	409	46	–*	33,702	1

Table 2: Computational results

Table 2 reveals that test instance 5–8 cannot be solved with the IP. A closer look into the Java console output revealed that CPLEX does not start preprocessing. Interestingly, for test instance 3 the optimal solution is already obtained by the therapist scheduling problem (ThSP) and hence this problem does not contain any cuts added by our cutting plane algorithm. However, the corresponding full problem formulation using the IP requires 1.25 hours to verify the optimal solution. Thus, the hospital cannot use the full IP formulation to solve their instances neither to optimality nor does the IP provide an initial solution. In contrast, using the CP algorithm, the hospital can solve their test instances to optimality.

6.4 Evaluation of lateness and unscheduled jobs

Table 3 provides the lateness as well as the total number of unscheduled jobs for all methods. Instead of reporting the original objective function (6) which minimizes the sum of the start times $\sum_i \sum_m \sum_t t \cdot x_{i,m,t}$ we report the sum of the lateness $\sum_i \sum_m \sum_t (t \cdot x_{i,m,t}) - E_i^{\text{job}}$. Obviously the two objectives lead to the same optimal solutions since the lateness objective is obtained from the start time objective by subtracting the constant term $\sum_i E_i^{\text{job}}$. We report the objective function performance in terms of lateness because the numbers are smaller. For the SAH we penalize non-feasible solutions by multiplying the number of unscheduled jobs by the length of the planning horizon T.

	Lateness				# Unscheduled Jobs			
Instance	IP	CP	SAH	ThSP	IP	CP	SAH	ThSP
1	2	2	6	2	0	0	0	0
2	5	5	55	5	0	0	2	0
3	9	9	33	9	0	0	1	0
4	20	20	34	7	0	0	1	0
5	–*	31	34	26	–*	0	1	0
6	–*	73	86	45	–*	0	0	0
7	–*	54	95	32	–*	0	3	0
8	–*	44	59	38	–*	0	1	0

Table 3: Evaluation of lateness and unscheduled jobs

Table 3 shows that the IP and the cutting plane algorithm schedule all jobs, which is not surprising because of constraints (6). Moreover, one can observe that the number of unscheduled jobs by SAH is between 0 and 3 across all test instances. As a consequence, no more than 6.67% of the total number of jobs remain unscheduled in any of the test instances run by the heuristic. In cases when SAH fails to schedule all jobs, the hospital may increase the duration of the therapists' shifts, cancel jobs or reschedule them.

7 Discussion, Limitations and Conclusions

7.1 Discussion and Limitations

The results have demonstrated that the cutting plane algorithm for scheduling therapy jobs and routing therapists hospital-wide can substantially reduce solution times. We argue that the approaches can be generalized for solving other problem instances in other service industries such as parcel delivery in which delivery locations are flexible (Cepolina and Farina (2015)).

Our paper does not provide a GUI-based decision support system which is directly applicable in the hospital. For a pathway towards the applicability in the hospital, our approaches have to be integrated in the hospital information system. The integration requires the development of a GUI-based front-end as well as the connection to a communication server which pulls the patient-information from the clinical information system.

Our approach does not address uncertainty. We outline below why this is appropriate for the therapist scheduling and routing problem and how we can utilize our deterministic approach in the case of a change of data. However, for related applications with considerable uncertainty of data as in surgery scheduling (Cardoen et al. (2010)), a deterministic approach is not suited any more and a stochastic approach has to be undertaken. Hence, future research should augment our deterministic approach to the stochastic case.

Since our model focuses on inpatients and scheduled outpatient jobs, we assume that inpatients are available at the ward. Therefore, those jobs shall be considered as a deterministic parameter. The hospital does not have a walk-in policy. Accordingly, on the day under consideration, it is unrealistic that new jobs have to be scheduled during the execution of the schedule. The only uncertainty with scheduled outpatient jobs is, however, no-shows which would cause idle time for the assigned therapist. Since utilization is not our objective, outpatient no-shows are not relevant for our problem.

In contrast to other health care operations such as e.g. surgeries (Cardoen et al. (2010)), the duration of therapy jobs is deterministic. Hence, we do not have to consider stochastic job durations.

When generating the schedule, therapists are assumed to be available. In the case of an unforeseen non-availability of a therapist (because of illness, weather conditions etc.), a new schedule has to be generated based on the remaining therapists. One way of obtaining feasible solutions would be canceling some jobs. The objective function would then have to maximize the number of jobs to be scheduled and some measure of deviation between the new and the former schedule.

7.2 Conclusions

In this paper, we addressed the problem of scheduling physical therapies and routing physical therapists hospital-wide. We have shown that small instances can be solved to optimality using a Mixed-Integer Program. Larger instances, however, can be solved to optimality within an acceptable solution time using a cutting plane method. The results of a heuristic algorithm that we proposed and which can be used for large-sized real world instances revealed a good trade-off between the objective to schedule therapy jobs early and the number of unscheduled jobs.

A Abbreviations, sets, indices and decision variables

Abbreviations

CP	Cutting plane
ICU	Intensive care unit
IP	Integer Program
MIP	Mixed-Integer Program
MMRCPSP	Multi-mode resource-constrained project scheduling problem
PT	Physical therapy
SAH	Sequential allocation heuristic
TC	Therapy center
ThSP	Therapist scheduling problem
ThSRP	Therapist scheduling and routing problem

Sets and indices

$c_{r,s}$	Duration to walk between room $r \in \mathcal{R}$ and room $s \in \mathcal{R}$
d_i^{job}	Duration of job $i \in \mathcal{I}$
$d_{i,j}^{\min}$	Minimum time lag referring to precedence relation $(i, j) \in \mathcal{E}$
Δ^{maxW}	Maximum total working time of a therapist
Δ^{regW}	Regular working time of a therapist; An extension requires a second break.
$\mathcal{E}$	Set of precedence relations
E_i^{job}	Earliest start of job $i \in \mathcal{I}$
$i_{1,p}^{\text{br}}$	First break of therapist p
$i_{2,p}^{\text{br}}$	Second break of therapist p
$\mathcal{I}$	Set of jobs
$\mathcal{I}^{\text{icu}} \subset \mathcal{I}$	Subset of ICU jobs
$\mathcal{I}^{\text{ward}} \subset \mathcal{I}$	Subset of ward jobs
$\mathcal{I}^{\text{tc}} \subset \mathcal{I}$	Subset of therapy center jobs
$\mathcal{I}_0$	Set of jobs including a dummy job that is carried out at the therapy center
K_r^{room}	Capacity of room $r \in \mathcal{R}$

$\mathcal{L}$	Set of break numbers
$\mathcal{L}_p$	Route of therapist $p \in \mathcal{P}$
$M_{i,j}$	A large constant dependent on combinations of jobs $i \in \mathcal{I}$ and $j \in \mathcal{I}$
$\mathcal{M}_i$	Set of modes for job $i \in \mathcal{I}$
$\mathcal{P}$	Set of therapists
L_i^{job}	Latest start of job $i \in \mathcal{I}$
$r_{i,m,p}^{\text{th}}$	1, if job $i \in \mathcal{I}$ requires therapist $p \in \mathcal{P}$ in mode $m \in \mathcal{M}_i$, 0 otherwise
$r_{i,m,r}^{\text{room}}$	1, if job $i \in \mathcal{I}$ requires room $r \in \mathcal{R}$ in mode $m \in \mathcal{M}_i$, 0 otherwise
$\mathcal{R}$	Set of rooms
$\mathcal{R}^{\text{icu}} \subset \mathcal{R}$	Subset of ICU rooms
$\mathcal{R}^{\text{ward}} \subset \mathcal{R}$	Subset of ward rooms
$\mathcal{R}^{\text{tc}} \subset \mathcal{R}$	Subset of therapy center rooms
$\mathcal{S}_p^{\text{e}}$	Set of times which denote the shift end of therapist p
S_i	Start time of therapy job $i \in \mathcal{I}$
S_i^*	Optimal start time of job $i \in \mathcal{I}$ as obtained by the full mathematical model
S_p^{sh}	Shift start of therapist p
T	Planning horizon
$\mathcal{T}$	Set of times
$\mathcal{W}_i$	Set of times for scheduling the start of job $i \in \mathcal{I}$
$\mathcal{X}^*$	Solutions of the master problem

Decision variables and objective function

$n_{i,j,p,r,s}$	1, if job $i \in \mathcal{I}$ preceeds job $j \in \mathcal{I}$ and are both assigned to the same therapist $p \in \mathcal{P}$ who executes job i in room $r \in \mathcal{R}$ and job j in room $s \in \mathcal{R}$, 0 otherwise
$w_{i,j,p,r,s}$	1, if job $i \in \mathcal{I}$ and job $j \in \mathcal{I}$ are assigned to the same therapist $p \in \mathcal{P}$ who executes job i in room $r \in \mathcal{R}$ and job j in room $s \in \mathcal{R}$, 0 otherwise
$x_{i,m,t}$	1, if job $i \in \mathcal{I}$ is started in mode $m \in \mathcal{M}_i$ at time $t \in \mathcal{W}_i$, 0 otherwise
$y_{p,t}$	1, if therapist $p \in \mathcal{P}$ works at time $t \in \mathcal{T}$, 0 otherwise
z	Objective function
$z_{i,j}$	1, if job i is followed by job j, 0 otherwise

Table 4: Abbreviations, sets, indices and decision variables

B Linearization of the sequencing constraints

The sequencing constraints (15) contain a quadratic term $z_{i,j} \cdot w_{i,j,p,r,s}$ which we can linearize using standard methods (Nemhauser and Wolsey (1999)): We incorporate additional decision variables $n_{i,j,p,r,s}$ into the model. The variables are 1 if job i is performed in room r by therapist p followed by job j which is performed in room s. We replace constraints (15) by constraints (18)

$$\sum_{m \in \mathcal{M}_j} \sum_{t \in \mathcal{W}_j} t \cdot x_{j,m,t} - \left(\sum_{m \in \mathcal{M}_i} \sum_{t \in \mathcal{W}_i} t \cdot x_{i,m,t} + c_{r,s} + d_i^{\text{job}} \right)$$
$$\geq (n_{i,j,p,r,s} - 1) \cdot M_{i,j} \qquad \forall i \in \mathcal{I}, j \in \mathcal{I} \setminus i, p \in \mathcal{P}, r \in \mathcal{R}, s \in \mathcal{R} \setminus r \tag{18}$$

Moreover, we add constraints (19)–(21).

$$- z_{i,j} + n_{i,j,p,r,s} \leq 0 \qquad \forall i \in \mathcal{I}, j \in \mathcal{I} \setminus i, p \in \mathcal{P}, r \in \mathcal{R}, s \in \mathcal{R} \setminus r \tag{19}$$

$$- w_{i,j,p,r,s} + n_{i,j,p,r,s} \leq 0 \qquad \forall i \in \mathcal{I}, j \in \mathcal{I} \setminus i, p \in \mathcal{P}, r \in \mathcal{R}, s \in \mathcal{R} \setminus r \tag{20}$$

$$z_{i,j} + w_{i,j,p,r,s} - n_{i,j,p,r,s} \leq 1 \qquad \forall i \in \mathcal{I}, j \in \mathcal{I} \setminus i, p \in \mathcal{P}, r \in \mathcal{R}, s \in \mathcal{R} \setminus r \tag{21}$$

C Algorithms

C.1 Job-room-start time tuple generation

Algorithm 1 Algorithm for generating job-room-start time tuple

Set $\mathcal{L}_p = \{\emptyset\}$
for all $t \in \mathcal{T}$ **do**
    for all $i \in \mathcal{I}$ **do**
        if $t \geq E_i^{\text{job}}$ and $t \leq L_i^{\text{job}}$ **then**
            for all $m \in \mathcal{M}_i$ **do**
                for all $r \in \mathcal{R} : r_{i,m,r}^{room} = 1$ **do**
                    if $r_{i,m,p}^{\text{th}} = 1$ and $x_{i,m,t} = 1$ **then**
                        $\mathcal{L}_p \cup (i, r, t)$
                    end if
                end for
            end for
        end if
    end for
end for

C.2 Stage 1 of the SAH

Algorithm 2 Stage 1 of the sequential allocation heuristic

Initialize $x_{i,m,t} := 0 \quad \forall i \in \mathcal{I}$.
for all $i \in \mathcal{I} : E_i^{\text{job}} = L_i^{\text{job}}$ (in the order of E_i^{job}) **do**
    for all $t \in \mathcal{T}$ **do** (increasing order)
        if $t = E_i^{\text{job}}$ **then**
            for all $m \in \mathcal{M}_i$ **do**
                for all $p \in \mathcal{P} : r_{i,m,p}^{\text{th}} = 1$ (in increasing shift start order) **do**
                    for all $r \in \mathcal{R} : r_{i,m,r}^{room} = 1$ (increasing order) **do**
                        if $\forall \tau : t \leq \tau \leq t + d_i^{\text{job}} \wedge R_{k,\tau} \geq 1 \wedge R_{r,\tau} \geq 1$ **then**
                            $x_{i,m,t} := 1$;
                            $R_{k,\tau} := R_{k,\tau} - 1 \quad \forall t \leq \tau \leq t + d_i^{\text{job}}$;
                            $R_{r,\tau} := R_{r,\tau} - 1 \quad \forall t \leq \tau \leq td_i^{\text{job}}$;
                        end if
                    end for
                end for
            end for
            if $x_{i,m,t} := 1$; **then**
                break;
            end if
        end if
    end for
end for

C.3 Stage 2 of the SAH

Algorithm 3 Stage 2 of the sequential allocation heuristic

for all $i \in \mathcal{I} : E_i^{\text{job}} \neq L_i^{\text{job}}$ (in increasing order of E_i^{job}) **do**
  for all $t \in \mathcal{T}$ **do** (therapists ordered in increasing shift start or increasing shift start)
    if $t \in \mathcal{W}_i$ **then**
      for all $m \in \mathcal{M}_i$ **do**
        for all $p \in \mathcal{P} : r_{i,m,p}^{\text{th}} = 1$ (in increasing shift start order) **do**
          for all $r \in \mathcal{R} : r_{i,m,r}^{room} = 1$ (increasing order) **do**
            $t := t + d_{min}$;
            if $\forall \tau : t \leq \tau \leq t + d_i^{\text{job}} + d_t \wedge R_{k,\tau} \geq 1 \wedge R_{r,\tau} \geq 1$ **then**
              $x_{i,m,t} := 1$;
              $R_{p,\tau} := R_{p,\tau} - 1 \quad \forall t \leq \tau \leq t + d_i^{\text{job}} + d_t$;
              $R_{k,\tau} := R_{k,\tau} - 1 \quad \forall t + d_t \leq \tau \leq t + d_i^{\text{job}} + d_t$;
            end if
          end for
        end for
        if $x_{i,m,t} := 1$ **then**
          break;
        end if
      end for
    end if
  end for
end for

References

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

A. Beaudry, G. Laporte, T. Melo and S. Nickel. Dynamic transportation of patients in hospitals. *OR Spectrum*, 32(1):77–107, 2010.

Bureau of Labor Statistics. Physical therapists. *U.S. Department of Labor, Occupational Outlook Handbook, 2014-15 Edition*, 2012. last accessed: October 5th, 2014.

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

E. Cepolina and A. Farina. A new urban freight distribution scheme and an optimization methodology for reducing its overall cost. *European Transport Research Review*, 7(1):1, 2015.

M. Cissé, S. Yalçındağ, Y. Kergosien, E. Şahin, C. Lenté and A. Matta. OR problems related to Home Health Care: A review of relevant routing and scheduling problems. *Operations Research for Health Care*, 13-14:1–22, 2017.

M. Delorme. *Mathematical Models and Decomposition Algorithms for Cutting and Packing Problems.* PhD thesis, University of Bologna, 2017.

S. H. H. Doulabi, L.-M. Rousseau and G. Pesant. A constraint-programming-based branch-and-price-and-cut approach for operating room planning and scheduling. *INFORMS Journal on Computing*, 28(3): 432–448, 2016.

M. Erhard, J. Schoenfelder, A. Fügener and J. O. Brunner. State of the art in physician scheduling. *European Journal of Operational Research*, 265(1):1–18, 2018.

C. Fikar and P. Hirsch. Home health care routing and scheduling: A review. *Computers & Operations Research*, 77:86–95, 2017.

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

J. Griffiths, J. Williams and R. Wood. Scheduling physiotherapy treatment in an inpatient setting. *Operations Research for Health Care*, 1(4):65–72, 2012.

T. Hanne, T. Melo and S. Nickel. Bringing robustness to patient flow management through optimized patient transports in hospitals. *Interfaces*, 39(3):241–255, 2009.

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

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

K. H. Kang, B. K. Lee, Y. H. Lee and Y. H. Lee. A heuristic for the vehicle routing problem with due times. *Computers & Industrial Engineering*, 54(3):421–431, 2008.

R. Kolisch and A. Drexl. Local search for nonpreemptive multi–mode resource–constrained project scheduling. *IIE Transactions*, 29(11):987–999, 1997.

O. Koné, C. Artigues, P. Lopez and M. Mongeau. Event-based MILP models for resource-constrained project scheduling problems. *Computers & Operations Research*, 38(1):3–13, 2011.

G. Laporte and F. V. Louveaux. The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13(3):133–142, 1993.

J. Marynissen and E. Demeulemeester. Literature review on integrated hospital scheduling problems. 2016.

G. L. Nemhauser and L. A. Wolsey. *Integer and Combinatorial Optimization*. Wiley, New York, 1999.

V. Roshanaei, C. Luong, D. M. Aleman and D. Urbach. Propagating logic-based benders' decomposition approaches for distributed operating room scheduling. *European Journal of Operational Research*, 257 (2):439–455, 2017a.

V. Roshanaei, C. Luong, D. M. Aleman and D. R. Urbach. Collaborative operating room planning and scheduling. *INFORMS Journal on Computing*, 29(3):558–580, 2017b.

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

V. Schmid and K. F. Doerner. Examination and operating room scheduling including optimization of intrahospital routing. *Transportation Science*, 48(1):59–77, 2014.

R. Stephan. Cardinality constrained combinatorial optimization: Complexity and polyhedra. *Discrete Optimization*, 7(3):99–113, 2010.

F. B. Talbot. Resource-constrained project scheduling with time-resource tradeoffs: The nonpreemptive case. *Management science*, 28(10):1197–1210, 1982.

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

D. L. Vetrano, F. Landi, S. L. D. Buyser, A. Carfì, G. Zuccalà, M. Petrovic, S. Volpato, A. Cherubini, A. Corsonello, R. Bernabei and G. Onder. Predictors of length of hospital stay among older adults admitted to acute care wards: a multicentre observational study. *European Journal of Internal Medicine*, 25(1):56–62, 2014.

B. Vieira, E. W. Hans, C. van Vliet-Vroegindeweij, J. van de Kamer and W. van Harten. Operations research for resource planning and-use in radiotherapy: a literature review. *BMC medical informatics and decision making*, 16(1):149, 2016.