

Decision support for rehabilitation hospital scheduling

Katja Schimmelpfeng^a, Stefan Helber^b and Steffen Kasper^b

^aChair of
Accounting and Control
Brandenburg University of Technology, Germany
Katja.Schimmelpfeng@tu-cottbus.de

^bDepartment of
Production Management
Leibniz Universität Hannover, Germany

October 13, 2010, revised March 9, 2011

Abstract

We present a detailed analysis of the patient and resource scheduling problem in rehabilitation hospitals. In practice, the predominantly therapeutical treatments and activities which are prescribed for the patients are typically scheduled manually. This leads to rigid and inefficient schedules which can have negative effects on the quality of care and the patients' satisfaction. We outline the conceptual framework of a decision support system for the scheduling process that is based on formal optimization models. To this end, we first develop a large-scale monolithic optimization model. Then we derive a numerically tractable hierarchical model system in order to deal with problem instances of realistic sizes. We report numerical results with respect to solution times, model sizes and solution quality.

1 Introduction

Rehabilitation hospitals serve to maintain and recover patients' health, often following surgery or other treatments in an acute hospital. Many patients remain in rehabilitation (rehab) hospitals for several weeks. Each day they can receive multiple treatments like massages or physiotherapy based on a physician's prescription and the available resources. Very often hundreds of different treatments are offered by a particular rehab hospital, depending on its individual specialization. As societies in many developed countries tend to age on average, the demand for rehabilitation will increase. Rehab hospitals are therefore an important part of the health care system because of their function, number, size and cost. In Germany, for example, more than 8 billion euros were spent on the 1000+ rehab hospitals in 2008 according to the German federal office of statistics, see Statistisches Bundesamt Deutschland (2010).

The predominantly therapeutic processes in these hospitals are typically personnel-intensive and very numerous. They usually have a fixed duration and are almost entirely planned in advance. For this reason, the quality of the schedules is highly important for the medical effectiveness and economic efficiency of a rehab hospital. The schedule for a large rehab hospital can easily comprise more than ten thousand appointments for hundreds of patients over several weeks using hundreds of different resources. Scheduling all of these activities is highly complex, time-consuming and very cumbersome. In practice, at best, database systems are used to retrieve and store schedule-related data. Even paper-and-pencil "solutions" can be found, in particular in smaller institutions. In any case, the scheduling task itself is usually performed manually, on a patient-by-patient basis and often continuously over time. This scheduling effort is often so tremendous that several (low-level) clerks are required to find "empty time slots" in the resources' calendar in order to "book in" an appointment for a patient. This leads to very rigid schedules that are difficult to change if indicated due to changes in a patient's health status and/or rehabilitation success. Furthermore, it is very difficult for the management to efficiently control the rehab processes in the short term, for example by reallocating resources, using overtime etc.

To overcome this problem, we develop in this paper formal mixed-integer linear programs (MILPs) to determine appointments for patients of rehab hospitals, subject to numerous constraints that are often found in practice. We show how these MILPs can be solved quickly using commercial state-of-the-art solvers. Our analysis of the scheduling problem shows that a well-designed decision support system based on such formal optimization models can be used to automatize much of the scheduling effort. As a result, the management of a rehab hospital can explicitly consider different schedules and hence think systematically in terms of alternatives. Furthermore, it becomes possible to address groups of patients differently, e.g., based on their economic attractiveness for the rehab hospital.

The remainder of the paper is organized as follows: In section 2 we analyze the scheduling problem in detail and comment on the related literature. This leads to several design considerations for our formal decision support system that we discuss in section 3. On this basis we first present a monolithic model that captures the (in our eyes) most crucial aspects of the considered scheduling problem. However, this monolithic model cannot be solved for medium-sized to large problem instances with the currently available optimization technology. For this reason, we decompose the monolithic model into a three-stage hierarchical system of optimization models to make the planning task numerically tractable. In section 4, we compare both approaches with respect to computation times, model sizes and the losses in solution quality due to the hierarchical decomposition. In section 5 we discuss directions for further research.

2 Problem statement and literature overview

2.1 Characteristics of rehabilitation hospitals

2.1.1 Patient types and revenue generation

From a scheduling point of view the patients of a rehab hospital can be very heterogenous. This can affect the way revenues for the rehab hospital are generated and the way patients have to be treated—medically as well as organizationally. In Germany, for example, rehab patients will typically have health insurance from either a public or a private insurance provider or be self-pay patients. Treating self-pay patients and those with private insurance typically leads to higher per-capita revenues for the rehab hospital than treating the first and largest group with public health insurance.

The management of a rehab hospital will usually strive to attract and satisfy as many patients from the highly profitable patient groups as possible. To this end, it can be very attractive for the management to have the ability to grant the economically attractive patient groups reliable schedules that reflect the temporal preferences of this (usually small) group. While the ethical aspects of this kind of prioritization are complex (and beyond the scope of this paper), the models presented below can at least make the underlying trade-offs transparent within a structured decision support system.

In addition to patients who permanently stay in the rehabilitation hospital for (usually) some weeks, other patients use only the hospital’s treatment facilities, but live outside in their homes or in hotels. For those outpatients it is very important to have reliable schedules in order to coordinate their treatment in the hospital with their outside activities. Inpatients, on the other hand, are in principle permanently available for treatment and can be informed about their personal schedule at short notice, e.g., a day or two in advance.

2.1.2 Prescriptions, treatments and activities

After the initial admission diagnosis, a physician typically prescribes a program of predominantly therapeutic treatments and activities for each patient. This specifies the type, frequency, and, possibly, time windows for each type of treatment. This program can follow a specific clinical or medical “pathway”, i.e., a standardized process plan for a given diagnosis. However, there is room for substantial variability, both with respect to the type and the number of the prescriptions. The type (or nature) of the prescriptions reflects the profile and philosophy of the particular hospital. The (possibly inflated) number of the prescribed replications of any of these prescriptions can reflect the physician’s individual experience with the hospital’s operational performance in the past.

The majority of the activities and treatments in a rehabilitation hospital is provided by non-physician health care professionals like physiotherapists, masseurs, nutritionists or balneotherapists. These treatments aim at regaining mobility, physical strength and emotional stability. Physicians provide diagnostic services and prescribe both kinds of activities. Quite frequently these activities have to be repeated several times over the length of the patient’s stay. The prescriptions of treatments and therefore also the schedule may have to be modified over the course of the therapy, given the development of the patient’s health status. Treatments which may not be scheduled on the same day to avoid harming the patient are listed in a conflict group.

Individual treatments address a single patient and his particular needs, for example in a physiotherapy. From the hospital’s point of view these treatments are expensive and require a substantial and individual scheduling effort. Other activities like cardiovascular training or aquarobics tend to be organized as group activities and are hence substantially cheaper from the hospital’s perspective. Collective activities often have to follow stable schedules as they affect the individual schedules of many patients and therapists.

2.1.3 Resources and processes

The resources and the processes in a rehabilitation hospital have specific features that differ substantially from those in acute hospitals. The therapeutic treatments tend to be personnel-intensive and very often they require active participation of the patient.

In acute hospitals, the operating room is typically the central and also a very expensive resource. Its schedule influences many other pre- and post-operation processes, see, e.g., Beliën et al. (2007, 2009), Guerriero and Guido (2011). In a rehab hospital, however, it is neither possible nor necessary to derive schedules for the entire institution from the schedule for a single type of bottleneck resource. In addition, the processes in a rehab hospital are plannable to a high degree as emergencies are only dealt with in acute hospitals. Furthermore, a therapeutic treatment (e.g., a 20-minute massage) is very often *defined* by its processing time so that it can be treated as deterministic for scheduling purposes.

It may be necessary for some treatments to ensure that only a specific resource conducts a series of identical or related treatments for a given patient. Consider the case of a particular psychotherapist who has to get to know the patient and to establish a relationship based on trust over a series of meetings. In these cases, resource-loyal scheduling may be required. In addition, it can be important to schedule collective activities so that groups with a stable membership of patients emerge, for example, in a psychologic group therapy.

2.1.4 Organizational structure

Rehab hospitals are very often specialized with respect to the medical services they provide and the patient types they serve. Physicians typically constitute a relatively small fraction of the workforce in rehab hospitals. Unlike in acute hospitals, they are often not assigned to a specific department, but perform a central function and hence prescribe treatments for the entire hospital. In this organizational setting, it is possible and actually quite common to centralize the scheduling function for the entire rehab hospital. Note that the situation in acute hospital differs substantially.

2.2 Established planning processes in rehabilitation hospitals

Established “planning methods” in practice typically focus on the manual management of calendars for all relevant resources such as therapists, rooms or equipment. For this purpose, either electronic database systems, physical scheduling boards or “paper-and-pencil solutions” are used.

Electronic database systems frequently model time in a discrete fashion using time slots with a length of five minutes. Clerks “book in” appointments for patients based on the prescriptions and, sometimes, patients’ wishes. This booking is performed in a sequential one-by-one manner. In large institutions, these clerks may even work spatially distributed. Very often, the schedule for a patient is (at least temporarily) fixed after the prescriptions have been booked in. This leads to schedules that can be very rigid, i.e., hard to adapt to changing patient’s needs over the course of the therapy or changes of the resource availability. As with the initial scheduling, any re-scheduling is done manually. For this reason, the scheduling process itself is expensive and time-consuming. It can easily lead to schedules that cause an inefficient use of the resources.

In addition, hospitals do not have the ability to automatically treat their different patient or “customer” segments individually in order to optimize the profit of the rehab hospital or the resource utilization. The hospital management might want to gain the ability to offer schedules to the profitable private health care and self-pay patients which reflect the patient’s individual temporal wishes. The remaining majority of the public health care patients could then be automatically scheduled “around” the exogenously

fixed appointments for the more profitable patients. In currently available systems, this is achieved via manual (re-)scheduling.

Commercial scheduling systems, to the best of our knowledge, tend (at best) to offer undocumented scheduling algorithms to support the booking process described above. These algorithms often operate directly (and hence slowly) on the mass storage holding the database, nowadays often using client-server architectures.

2.3 Literature overview

Decades of research on health care operations management and research of health care systems have led to numerous publications which address diverse topics such as facility location, capacity planning, shift scheduling, inventory management, to mention only a few. Due to the interdisciplinary nature of the topic, these publications are distributed over different academic disciplines from medicine to mathematics. The recent handbooks by Bradeau et al. (2005) and Hall (2006) provide a good overview of the different fields. A detailed “ORchestra Bibliography” is provided by the *Centre for Healthcare Operations Improvement & Research (CHOIR)* at the University of Twente (www.utwente.nl/choir). The websites of the “Centre for Research in Healthcare Engineering” at the University of Toronto (crhe.mie.utoronto.ca) as well as the site by Franklin Dexter (www.franklindexter.net) also provide many valuable references to research papers.

There is a broad body of literature on scheduling in (acute) hospitals. Much of this literature focuses on operating room scheduling. A recent review is given by Cardoen et al. (2010). Given the immense importance of acute hospitals and the central role of the operating room, this scheduling task will continue to attract a lot of research. Other publications address the efficient usage of expensive diagnostic equipment, see, e.g., Kolisch and Sickinger (2008) and Sickinger (2008), dealing with radiologic diagnostic services. A second important branch of research addresses nurse scheduling. Cheang et al. (2003) provided a bibliographic survey and Kellogg and Walczak (2007) discussed why much of this research is never implemented and how systems used in practice are actually developed. Scheduling physicians has also attracted a lot of research, see, for example, Brunner et al. (2009) as well as Brunner (2010), and the references given therein.

It is interesting to note that, compared to the substantial number of publications dealing with planning and scheduling in acute hospitals, only a few papers address the scheduling problems in rehabilitation hospitals. In an early work, Podgorelec and Kokol (1997) described in very general and non-technical terms a genetic algorithm for patient scheduling to a limited number of human or technical therapeutic resources. Later Chien et al. (2008) developed a genetic algorithm to sequence (for a given day) the patients requiring rehabilitation treatment. They use a hybrid job scheduling model that aims at balancing the makespan against the waiting time of patients. This research does not address the problem of how to derive a schedule for a complete rehab hospital over a

planning horizon of several weeks. A genetic algorithm is also used in Chien et al. (2009) in combination with data mining to solve a physical therapy scheduling problem. Ogulata et al. (2008) developed a three-stage decision support system based on algebraic modeling for the physiotherapy and rehabilitation services provided by the Cukurova University Hospital in Turkey. At the first stage, a decision is made as to whether to accept or to refuse a patient for treatment in the department. The second stage assigns patients to physiotherapists and the third determines the schedule by assigning the treatments for a given day to a particular time slot.

A common feature of these (rehabilitation-oriented) papers is that they tend to focus on a single department. The rare opposite case is studied by Vanberkel et al. (2010) who presented “A survey of health care models that encompass multiple departments”, i.e., they discussed models that aim to coordinate the schedule for the entire clinic. However, they concentrated on the situation in acute hospitals that face emergencies, that have the operating room as a central and dominating resource and that exhibit decentralized scheduling functions in the different medical departments. From our analysis of the available literature, we conclude that software vendors which develop information and decision support systems for rehabilitation hospitals are currently offered very limited support from the operations research literature. Our organizational and operational analysis of the planning and scheduling problems furthermore suggests that these problems should be easier to solve than those in acute hospitals, given the much more stable information basis in rehab hospitals. We therefore try to close this gap by developing a general and conceptual framework for model-based decision support systems for rehab hospitals that reflects the needs described in subsection 2.1.

3 Modeling approaches

3.1 Design considerations for a decision support system

From our discussions with managers and planners in rehab hospitals and with software vendors we conclude that any decision support system which is actually useful in practice must permit a substantial degree of manual scheduling, see Kellogg and Walczak (2007). Planners in rehab hospitals tend to schedule the often periodic collective activities manually with respect to time and resource allocation. In addition, they may need to be able to exogenously fix some appointments, e.g., for self-pay patients or out patients, while leaving it to the systems to schedule the remaining (not fixed) activities or treatments automatically. Furthermore, it may be necessary to permanently assign a particular resource to a patient-treatment combination to ensure resource loyalty. It may also be necessary to assign a patient to a particular group for a given collective activity in order to achieve group loyalty.

We aim at a decision support system with the ability to automatically schedule activi-

ties for patients and to assign the required resources, subject to a number of rehabilitation-specific constraints. From a mathematical point of view this leads to an assignment problem. The assumed objective is to schedule as many of the prescribed activities as possible, given the available resources and the exogenous scheduling decisions about, for example, group activities. Note that a naive use of such a system can lead to schedules that are unacceptable from the perspective of an individual patient or therapist. Consider, e.g., the case that the prescribed treatments for an individual patient are not scheduled to a sufficient degree because of resource shortage. In addition, the workload assigned to individual therapists may also vary substantially and hence also raise debates about fairness. In order to deal with issues of this type, it is important that a decision support system can quickly generate alternative schedules. Then it can be used iteratively until exogenous decisions (by the planner) about resource allocations are found that lead to satisfying endogenous decisions about the schedule. Such a decision support system will merely help the planner to develop satisfying schedules and not take over this function.

We envision using the decision support system in a rolling fashion. Each day, the planners could solve the models to determine fixed (or “frozen”) schedules for the next few days and tentative schedules for the remaining days of the planning horizon. Only fixed schedules would be communicated to patients and therapists. This way planners would preserve their flexibility to react to parameter changes due to new prescriptions, unavailability of therapists etc. Note that over the planning horizon, some patients leave the hospital while others arrive for which the precise prescriptions are not known. However, based on their illness and/or prior treatment in an acute hospital, the planners could use standardized clinical pathways to roughly determine the resource requirements over time for a particular future patient and plan based on preliminary or projected prescription that would be updated after the patient’s initial examination. In such an environment a hierarchical planning system seems to be particularly adequate that treats the near future at a greater level of details than the more distant future.

In this daily planning process, the planner would check both the utilization of the resources and the number of prescribed treatments that could not be (at least tentatively) scheduled. If the resulting plan were unacceptable, parameters would be changed and the model(s) would be solved again to determine a plan that is acceptable from both the patient’s and the planner’s perspective.

The system should be completely based on algebraic models, i.e., it should not use problem-specific algorithms. The conceptual framework presented in this paper can therefore easily be modified to consider additional aspects of the planning problem that may be relevant in a particular rehab hospital.

3.2 General assumptions and notation

The major assumptions of our models can be separated into four classes: assumptions concerning time, resources, patients, and activities/treatments.

Time

- T1 Time is divided into discrete periods consisting of days d (Monday to Friday) and intra-day slots t (e.g., 5-minute intervals).
- T2 Transition times are not considered.
- T3 Activities require resources for the whole duration of a treatment. The duration is given by the number s_m of consecutive intra-day time slots for treatment m .
- T4 Patients need to rest after treatment m for ps_m time slots.

Resources

- R1 Individual resources f like therapists, medical devices, and rooms are grouped into homogeneous resource categories \hat{f} .
- R2 A binary parameter $av_{fdt} = 1$ indicates that resource f is available at time slot (d, t) , it is 0 otherwise.
- R3 Treatment m requires $u_{\hat{f}m}$ units of resource category \hat{f} simultaneously.

Patients

- P1 A number of r_{pm} replications of treatment m is prescribed for patient p .
- P2 Patient p is available for treatments at days $d \in D_p$.
- P3 Each patient p has an individual stress limit a_p per day. A single replication of treatment m places b_m stress units on the patient on that day.
- P4 At any moment in time, a patient p can attend at most one treatment.

Activities/treatments

- A1 Treatment m is provided for either a single patient p or a group q of up to k_m patients simultaneously.
- A2 Patient p receiving a collective treatment can be exogenously assigned to a particular group q and scheduled to time slot (d, t) to ensure group and/or resource loyalty.
- A3 Group q for a collective activity/treatment is exogenously assigned to each required resource f and (exogenously) scheduled to time slot (d, t) .

A4 An individual treatment of patient p can be exogenously assigned to each required resource f and (exogenously) scheduled to time slot (d, t) .

A5 Any treatment m is provided at most once per day for each patient p .

A6 No two treatments m belonging to the same conflict group c may be scheduled for the same day for any patient p .

A7 Multiple replications r_m of treatment m for patient p prescribed for a time window $[fi_{pm}, la_{pm}]$ (in days) should be distributed (roughly) evenly over the days of this time window.

A8 If two different treatments m_1 and m_2 for a patient p are assigned to the same day d , it may be necessary that m_1 precedes m_2 .

The following general notation is used in both the monolithic and the hierarchical model system(s):

(Ordered) sets and indices

$c \in C$	conflict groups to avoid scheduling conflicting treatments on the same day
$d, \delta \in D$	days (e.g., Monday to Friday)
D_p	days patient p is available
$f \in F$	resources (i.e., specific therapists, medical devices, rooms)
$\hat{f} \in \hat{F}$	homogeneous resource categories
$F_{\hat{f}}$	individual resources f belonging to resource category \hat{f}
\hat{F}_m	resource categories required for treatment m
lM_{pf}	treatments for patient p which require resource loyalty and are exogenously allocated to resource f
lQ_{pm}	exogenously assigned group(s) for group-loyal treatment m of patient p
$m \in M$	treatments/activities
M_c	treatments belonging to conflict group c
$M_{\hat{f}}$	treatments which require resource category \hat{f}
M_p	treatments prescribed for patient p
N_{m_1}	treatments which may not be scheduled prior to treatment m_1 on a given day
$p \in P$	patients
P_m	patients who need treatment m
$q \in Q_m$	groups for treatment m
$t, \tau \in T$	intra-day time slots (e.g., 5-minute time slots)

Parameters

a_p	stress limit for patient p per day
b_m	stress factor of treatment m
$d_{m_1 m_2}$	minimal break (in 5-minute time slots) between treatments m_1 and m_2
fi_{pm}	first possible day for treatment m of patient p
k_m	maximum group size of collective activity/treatment m
la_{pm}	latest possible day for treatment m of patient p
ps_m	duration (of the break) to recover from treatment m (in time slots)
r_{pm}	number of prescribed replications of treatment m for patient p
s_m	duration of treatment m (in 5-minute time slots)
$u_{\hat{f}m}$	number of resource units of category \hat{f} needed for treatment m

3.3 Monolithic planning approach

3.3.1 Specific notation for the monolithic model

The following notation is specific to the monolithic model:

(Ordered) set

GM_{pdt} exogenously scheduled treatments m of patient p that start at time slot (d, t)

Parameters

av_{fdt}	equals 1, if resource f is available at time slot (d, t) , (0, otherwise)
w_{fqdt}	equals 1, if resource f is exogenously assigned to group q starting at time slot (d, t) , (0, otherwise)
z_{qdt}	equals 1, if group q is exogenously scheduled to start at time slot (d, t) , (0, otherwise)

Binary decision variables

v_{pqdt}	equals 1, if patient p is assigned to group q starting at time slot (d, t) , (0, otherwise)
x_{pmdt}	equals 1, if treatment m for patient p starts at time slot (d, t) , (0, otherwise)
y_{pmfdt}	equals 1, if (individual) treatment m of a patient p starting at time slot (d, t) uses resource f , (0, otherwise)

3.3.2 Monolithic model

The objective function (1) maximizes the number of treatments scheduled for all patients in the rehabilitation hospital, taking the possible time slots into account.

$$\text{Max } ZF = \sum_{p \in P} \sum_{m \in M_p} \sum_{\substack{d \in D_p \\ f_{ipm} \leq d \leq l_{apm}}} \sum_{\substack{t \in T \\ t \leq |T| - s_m + 1}} x_{pmdt} \quad (1)$$

The following group of constraints addresses the patients:

$$\sum_{m \in M_p} \sum_{\substack{\tau \in T \\ t - s_m - p s_m + 1 \leq \tau \leq t}} x_{pmd\tau} \leq 1 \quad p \in P, d \in D_p, t \in T \quad (2)$$

$$\sum_{\substack{d \in D_p \\ f_{ipm} \leq d \leq l_{apm}}} \sum_{\substack{t \in T \\ t \leq |T| - s_m + 1}} x_{pmdt} \leq r_{pm} \quad p \in P, m \in M_p \quad (3)$$

$$\sum_{m \in M_p} \sum_{\substack{t \in T \\ t \leq |T| - s_m + 1}} b_m \cdot x_{pmdt} \leq a_p \quad p \in P, d \in D_p \quad (4)$$

$$\sum_{\substack{\delta \in D_p \\ d \leq \delta \leq d + [\frac{l_{apm} - f_{ipm} + 1}{r_{pm}}] - 1}} \sum_{\substack{t \in T \\ t \leq |T| - s_m + 1}} x_{pmd\delta t} \leq 1 \quad p \in P, m \in M_p, \\ d \in D_p, f_{ipm} \leq d \leq l_{apm} \quad (5)$$

$$\begin{aligned} & \sum_{\substack{t \in T \\ t \leq |T| - s_{m_2} + 1}} t \cdot x_{pm_2 dt} - \sum_{\substack{t \in T \\ t \leq |T| - s_{m_1} + 1}} t \cdot x_{pm_1 dt} \\ & \geq s_{m_1} + d_{m_1 m_2} - |T| \cdot (1 - \sum_{\substack{t \in T \\ t \leq |T| - s_{m_1} + 1}} x_{pm_1 dt}) \\ & - 2 \cdot |T| \cdot (1 - \sum_{\substack{t \in T \\ t \leq |T| - s_{m_2} + 1}} x_{pm_2 dt}) \end{aligned} \quad p \in P, d \in D_p, \\ m_1, m_2 \in M_p, m_2 \in N_{m_1} \quad (6)$$

$$\sum_{m \in M_p \cap M_c} \sum_{\substack{t \in T \\ t \leq |T| - s_m + 1}} x_{pmdt} \leq 1 \quad p \in P, d \in D_p, c \in C \quad (7)$$

Restrictions (2) guarantee that each patient p can be scheduled for only one treatment m at time (d, t) . No more than the number of prescribed activities may be scheduled (3). Constraints (4) ensure that the total stress due to the treatments a patient p receives on day d does not exceed the individual stress limit a_p . The prescribed replications of treatment m have to be distributed (roughly) evenly within the given time window $[f_{ipm}, l_{apm}]$ for patient p , see restrictions (5). Restrictions (6) enforce a precedence relation between two treatments assigned to the same day. Restrictions (7) prohibit to schedule more than one treatment $m \in M_c$ of a conflict group c on any day d for any patient p . The next group of constraints deals with the resources:

$$\begin{aligned} \sum_{f \in F_{\hat{f}}} y_{pmfdt} &= u_{\hat{f}m} \cdot x_{pmdt} & p \in P, m \in M_p, k_m = 1, \\ && \hat{f} \in \hat{F}_m, d \in D_p, t \in T, t \leq |T| - s_m + 1 \\ && fi_{pm} \leq d \leq la_{pm} \end{aligned} \tag{8}$$

$$\begin{aligned} &\sum_{p \in P} \sum_{\substack{m \in M_p \cap M_{\hat{f}} \\ k_m = 1}} \sum_{\substack{\tau \in T \\ t - s_m + 1 \leq \tau \leq t}} y_{pmfd\tau} \\ &\leq av_{fdt} - \sum_{\substack{m \in M_{\hat{f}} \\ k_m > 1}} \sum_{q \in Q_m} \sum_{\substack{\tau \in T \\ t - s_m + 1 \leq \tau \leq t}} w_{fqd\tau} \quad \hat{f} \in \hat{F}, f \in F_{\hat{f}}, d \in D, t \in T \end{aligned} \tag{9}$$

Constraints (8) allocate the number $u_{\hat{f}m}$ of resources units of resource category \hat{f} which are required for individual treatments ($k_m = 1$) to the scheduled treatment m of patient p . Resource f can only be allocated to an individual treatment if it is available and not exogenously allocated to a collective/group activity, see constraints (9). The next two restrictions address collective activities ($k_m > 1$):

$$\begin{aligned} \sum_{q \in Q_m} z_{qdt} \cdot v_{pqdt} &= x_{pmdt} & p \in P, m \in M_p, k_m > 1, t \in T, \\ && d \in D_p, fi_{pm} \leq d \leq la_{pm}, t \leq |T| - s_m + 1 \end{aligned} \tag{10}$$

$$\begin{aligned} \sum_{p \in P_m} v_{pqdt} &\leq k_m \cdot z_{qdt} & m \in M, k_m > 1, q \in Q_m, d \in D, \\ && t \in T, t \leq |T| - s_m + 1 \end{aligned} \tag{11}$$

Restrictions (10) assign the patients requiring collective treatment m to a group q starting in time slot (d, t) that is dedicated to this treatment m . The group size k_m for treatment m must not be exceeded by any group q , see (11). The remaining three restrictions reflect exogenous settings with respect to patients, group loyalty and factor loyalty.

$$\begin{aligned} x_{pmdt} &= 1 & p \in P, d \in D_p, t \in T, m \in GM_{pdt}, \\ && fi_{pm} \leq d \leq la_{pm}, t \leq |T| - s_m + 1 \end{aligned} \tag{12}$$

$$\begin{aligned} v_{pqdt} &= 1 & p \in P, d \in D_p, t \in T, m \in GM_{pdt}, \\ && fi_{pm} \leq d \leq la_{pm}, q \in lQ_{pm}, t \leq |T| - s_m + 1 \end{aligned} \tag{13}$$

$$\begin{aligned} y_{pmfdt} &= x_{pmdt} & p \in P, f \in F, m \in lM_{pf}, k_m = 1, \\ && d \in D_p, fi_{pm} \leq d \leq la_{pm}, t \in T, t \leq |T| - s_m + 1 \end{aligned} \tag{14}$$

Exogenously scheduled treatments of patients have to be enforced, see (12). A possible exogenous assignment of patients to particular groups q is treated in (13). Equations (14) assure that for individual treatments in the set lM_{pf} that require resource loyalty, only the preselected resources f are assigned to the combination of patient and treatment. Variable definitions are given in section 3.3.1.

3.4 Hierarchical planning approach

3.4.1 Overview and problem decomposition

As our numerical experiments (see section 4) indicated that it is not possible to solve the monolithic model for medium-sized or large problem instances using current personal computers, we decomposed the model into a hierarchical three-stage model system:

- Stage 1: The Stage 1 model assigns the prescribed treatments for the patients to single days over the whole planning horizon (e.g., 30 days). Time is modeled at the aggregate level of days and capacity constraints are mainly considered at the aggregate level of resource categories \hat{f} .
- Stage 2: For each day d of the planning horizon we solve a first intra-day model. We consider the treatments assigned to that day within Stage 1 and schedule them to 5-minute time slots. Capacity constraints are mainly modeled at the aggregate level of resource categories \hat{f} .
- Stage 3: Given the intra-day temporal assignments of treatments to time slots for a given day d from Stage 2, we finally assign specific resources f to each scheduled individual treatment. We furthermore assign patients that require group treatments to the specific groups offering this treatment. The assignment for a particular combination of patient p and treatment m can be limited to a particular group q to ensure group loyalty.

As a result, we obtain a schedule that meets all the constraints of the monolithic model. However, it may contain fewer scheduled treatments than the solution of the monolithic model due to decomposition losses. Note that none of the models can ever be mathematically infeasible, but in an extreme case, it might (like the monolithic model) lead to a “solution” without any scheduled treatments. We discuss the decomposition losses in section 4. The three-stage model system is described in detail below together with the specific notation needed in addition to those given in section 3.2.

3.4.2 Specific notation for the hierarchical model system

The modified notation for the hierarchical model system can mainly be derived from those for the monolithic model by omitting the index for intra-day time slots t for parameters and decision variables of the Stage 1 model and by omitting the index d for days from the Stage 2 and 3 models. To avoid confusion with the notation of the monolithic model, we add superscripts in brackets to sets, parameters and decision variables that appear similarly in both modeling approaches. The superscripts indicate the stage that uses this notational element.

(Ordered) sets and indices

$GM_{pd}^{(1)}$	exogenously scheduled treatments m of patient p for day d (Stage 1)
$GM_{pt}^{(2,3)}$	exogenously scheduled treatments m of patient p to start at time t of the current day (Stages 2 and 3)
$l\hat{F} \subseteq \hat{F}$	subset of resource categories \hat{f} that may require resource-loyal scheduling
lQ_f	groups which require resource f
$M_p^{(2,3)}$	treatments for patient p assigned (in the Stage 1 solution) to the current day
$P_m^{(2,3)}$	patients with treatment m assigned (in the Stage 1 solution) to the current day

Parameters

$av_{ft}^{(2,3)}$	equals 1, if resource f is available at time slot t of the current day, and 0 otherwise (i.e., $av_{ft}^{(2,3)} = av_{fdt}$ of day d)
$cap\hat{F}_{\hat{f}d}$	available capacity (in hours) of resource category \hat{f} at day d (assuming 5-minute time slots, i.e., $cap\hat{F}_{\hat{f}d} = \sum_{f \in F_{\hat{f}}} \sum_{t \in T} \frac{av_{fdt}}{12}$ of day d)
$capF_{fd}$	available capacity (in hours) of resource f at day d (i.e., $capF_{fd} = \sum_{t \in T} \frac{av_{fdt}}{12}$ of day d)
n_{fm}	capacity requirement (in hours) of resource category \hat{f} for treatment m
$u_{ft}^{(2)}$	number of units of resource category \hat{f} available at time t of the current day d , (i.e., $u_{ft}^{(2)} = \sum_{f \in F_{\hat{f}}} av_{fdt}$ of day d)
$w_{fqt}^{(3)}$	equals 1, if resource f is exogenously assigned to group q starting at time t of the current day, and 0 otherwise (i.e., $w_{fqt}^{(3)} = w_{fqdt}$ of day d)
$z_{qd}^{(1)}$	equals 1, if group q is exogenously scheduled at day d , and 0 otherwise (i.e., $z_{qd}^{(1)} = \sum_{t \in T} z_{qdt}$ of day d)
$z_{qt}^{(2,3)}$	equals 1, if group q is exogenously scheduled to start at time t of the current day, and 0 otherwise (i.e., $z_{qt}^{(2,3)} = z_{qdt}$ of day d)

Binary decision variables

$x_{pmd}^{(1)}$	equals 1, if treatment m of patient p is scheduled at Stage 1 for day d , (0, otherwise)
$x_{pmt}^{(2)}$	equals 1, if treatment m of patient p is scheduled at Stage 2 to start at time t of the current day, (0, otherwise)
$x_{pmt}^{(3)}$	equals 1, if treatment m of patient p is scheduled at Stage 3 to start at time t of the current day, (0, otherwise)
$v_{pqt}^{(3)}$	equals 1, if patient p is assigned at Stage 3 to group q starting at time t , (0, otherwise)
$y_{pmft}^{(3)}$	equals 1, if (individual) treatment m of patient p starting at time t of the current day is assigned resource f at Stage 3, (0, otherwise)

3.4.3 Stage 1: Daily model

The objective function (15) maximizes the number of the tentatively scheduled treatments:

$$\text{Max } ZF = \sum_{p \in P} \sum_{m \in M_p} \sum_{\substack{d \in D_p \\ fi_{pm} \leq d \leq la_{pm}}} x_{pmd}^{(1)} \quad (15)$$

The following group of restrictions is the aggregated equivalent of constraints (3) to (7) from the monolithic model:

$$\sum_{\substack{d \in D_p \\ fi_{pm} \leq d \leq la_{pm}}} x_{pmd}^{(1)} \leq r_{pm} \quad p \in P, m \in M_p \quad (16)$$

$$\sum_{\substack{m \in M_p \\ fi_{pm} \leq d \leq la_{pm}}} b_m \cdot x_{pmd}^{(1)} \leq a_p \quad p \in P, d \in D_p \quad (17)$$

$$\sum_{\substack{\delta \in D_p \\ d \leq \delta \leq d + \lfloor \frac{la_{pm} - fi_{pm} + 1}{r_{pm}} \rfloor - 1}} x_{pmd}^{(1)} \leq 1 \quad p \in P, m \in M_p, d \in D_p, fi_{pm} \leq d \leq la_{pm} \quad (18)$$

$$\sum_{\substack{m \in M_p \cap M_c \\ fi_{pm} \leq d \leq la_{pm}}} x_{pmd}^{(1)} \leq 1 \quad p \in P, d \in D_p, c \in C \quad (19)$$

The next three constraints reflect the capacities of groups for collective treatments, resources at the category level and—where required for resource-loyal scheduling—at the level of individual resources:

$$\sum_{p \in P_m} x_{pmd}^{(1)} \leq k_m \cdot \sum_{q \in Q_m} z_{qd}^{(1)} \quad m \in M, k_m > 1, d \in D \quad (20)$$

$$\begin{aligned} & \sum_{p \in P} \sum_{\substack{m \in M_p \cap M_f \\ k_m=1}} n_{\hat{f}m} \cdot x_{pmd}^{(1)} \\ & \leq cap\hat{F}_{fd} - \sum_{\substack{m \in M_f \\ k_m>1}} \sum_{q \in Q_m} n_{\hat{f}m} \cdot z_{qd}^{(1)} \end{aligned} \quad \hat{f} \in \hat{F}, d \in D \quad (21)$$

$$\begin{aligned} & \sum_{p \in P} \sum_{\substack{m \in lM_{pf} \\ k_m=1}} \frac{n_{\hat{f}m}}{u_{\hat{f}m}} \cdot x_{pmd}^{(1)} \\ & \leq capF_{fd} - \sum_{\substack{m \in M \\ k_m>1}} \sum_{q \in lQ_f \cap Q_m} \frac{n_{\hat{f}m}}{u_{\hat{f}m}} \cdot z_{qd}^{(1)} \end{aligned} \quad \hat{f} \in l\hat{F}, f \in F_{\hat{f}}, d \in D \quad (22)$$

Constraints (20) ensure that aggregate capacity limits of groups providing collective treatments/activities are respected. The aggregated capacity limited of each resource category

\hat{f} must not be exceeded by the scheduled individual treatments and the usage of the resources for group treatments, see (21). In addition, an individual capacity constraint has to be met for each resource that is required for a treatment which requires resource loyalty, see (22). The last constraint addresses possible exogenous settings:

$$x_{pmd}^{(1)} = 1, \quad p \in P, d \in D_p, m \in GM_{pd}^{(1)} \quad (23)$$

The treatments scheduled exogenously to day d (collected in $GM_{pd}^{(1)}$) have to be enforced (23).

3.4.4 Stage 2: Time-oriented intra-day model

The objective function (24) maximizes the number of tentatively scheduled treatments of the patients for the current day, given the assignment of treatments $M_p^{(2,3)}$ to that day determined by the Stage 1 model:

$$\text{Max } ZF = \sum_{p \in P} \sum_{m \in M_p^{(2,3)}} \sum_{\substack{t \in T \\ t \leq |T| - s_m + 1}} x_{pmt}^{(2)} \quad (24)$$

Note that it is possible that a treatment is assigned to a particular day in the solution of the Stage 1 model, but cannot be scheduled in the solution of the Stage 2 model for that day, for example, as patients can receive only one treatment at a time.

Restrictions (25) are the counterparts of restrictions (2). They ensure that any patient attends at most one treatment at a time. Restrictions (26) guarantee that each treatment is scheduled at most once per day. Similarly to restrictions (6), restrictions (27) take the given order of treatments into account:

$$\sum_{m \in M_p^{(2,3)}} \sum_{\substack{\tau \in T \\ t - s_m - p s_m + 1 \leq \tau \leq t}} x_{pm\tau}^{(2)} \leq 1, \quad p \in P, t \in T \quad (25)$$

$$\sum_{\substack{t \in T \\ t \leq |T| - s_m + 1}} x_{pmt}^{(2)} \leq 1, \quad p \in P, m \in MP_p^{(2,3)} \quad (26)$$

$$\begin{aligned} & \sum_{\substack{t \in T \\ t \leq |T| - s_{m_2} + 1}} t \cdot x_{pm_2 t} - \sum_{\substack{t \in T \\ t \leq |T| - s_{m_1} + 1}} t \cdot x_{pm_1 t} \\ & \geq s_{m_1} + d_{m_1 m_2} - |T| \cdot (1 - \sum_{\substack{t \in T \\ t \leq |T| - s_{m_1} + 1}} x_{pm_1 t}) \\ & - 2 \cdot |T| \cdot (1 - \sum_{\substack{t \in T \\ t \leq |T| - s_{m_2} + 1}} x_{pm_2 t}) \end{aligned} \quad p \in P, m_1, m_2 \in M_p^{(2,3)}, m_2 \in N_{m_1} \quad (27)$$

The capacity restrictions of the groups dedicated to treatment m have to be respected:

$$\sum_{p \in P_m^{(2,3)}} x_{pmt}^{(2)} \leq k_m \cdot \sum_{q \in Q_m} z_{qt}^{(2,3)} \quad m \in M, k_m > 1, \quad t \in T, t \leq |T| - s_m + 1 \quad (28)$$

The capacity constraints of the resource categories have to be met, see (29). If a treatment-patient combination needs a loyal assignment to a particular resource f , the capacity constraint for this specific resource has to be met as well, see (30):

$$\begin{aligned} & \sum_{p \in P} \sum_{\substack{m \in M_p^{(2,3)} \cap M_{\hat{f}} \\ k_m=1}} \sum_{\substack{\tau \in T \\ t-s_m+1 \leq \tau \leq t}} u_{\hat{f}m} \cdot x_{pmt}^{(2)} \\ & \leq u_{\hat{f}t}^{(2)} - \sum_{\substack{m \in M_{\hat{f}} \\ k_m>1}} \sum_{q \in Q_m} \sum_{\substack{\tau \in T \\ t-s_m+1 \leq \tau \leq t}} u_{\hat{f}m} \cdot z_{q\tau}^{(2,3)} \quad \hat{f} \in \hat{F}, t \in T \end{aligned} \quad (29)$$

$$\begin{aligned} & \sum_{p \in P} \sum_{\substack{m \in M_p^{(2,3)} \cap lM_{pf} \\ k_m=1}} \sum_{\substack{\tau \in T \\ t-s_m+1 \leq \tau \leq t}} x_{pmt}^{(2)} \\ & \leq av_{ft}^{(2,3)} - \sum_{\substack{m \in M_{\hat{f}} \\ k_m>1}} \sum_{q \in lQ_f \cap Q_m} \sum_{\substack{\tau \in T \\ t-s_m+1 \leq \tau \leq t}} z_{q\tau}^{(2,3)} \quad \hat{f} \in l\hat{F}, f \in F_{\hat{f}}, t \in T \end{aligned} \quad (30)$$

Exogenously scheduled treatments have to be enforced:

$$x_{pmt}^{(2)} = 1 \quad p \in P, t \in T, m \in GM_{pt}^{(2,3)} \quad (31)$$

3.4.5 Stage 3: Resource-oriented intra-day model

Given the assignment of treatments to time slots for the current day from Stage 2, the Stage 3 model maximizes in the objective function (32) the number of treatments that are assigned to individual resources and hence are *finally* scheduled:

$$\text{Max } ZF = \sum_{p \in P} \sum_{m \in M_p^{(2,3)}} \sum_{\substack{t \in T \\ x_{pmt}^{(2)}=1}} x_{pmt}^{(3)} \quad (32)$$

Restrictions (33) guarantee that all the resources required for single treatments are provided:

$$\sum_{f \in F_{\hat{f}}} y_{pmft}^{(3)} = u_{\hat{f}m} \cdot x_{pmt}^{(3)} \quad p \in P, m \in M_p^{(2,3)}, k_m = 1, \quad \hat{f} \in \hat{F}_m, t \in T, x_{pmt}^{(2)} = 1 \quad (33)$$

Constraints (34) ensure that the capacity constraint of each individual resource is respected:

$$\begin{aligned}
& \sum_{p \in P} \sum_{\substack{m \in M_p^{(2,3)} \cap M_{\hat{f}} \\ k_m=1}} \sum_{\substack{\tau \in T \\ t-s_m+1 \leq \tau \leq t \\ x_{pm\tau}^{(2)}=1}} y_{pmf\tau}^{(3)} \\
& \leq av_{ft}^{(2,3)} - \sum_{\substack{m \in M_{\hat{f}} \\ k_m > 1}} \sum_{q \in Q_m \cap lQ_f} \sum_{\substack{\tau \in T \\ t-s_m+1 \leq \tau \leq t \\ z_{q\tau}^{(2,3)}=1}} w_{fq\tau}^{(2,3)} \quad \hat{f} \in \hat{F}, f \in F_{\hat{f}}, t \in T
\end{aligned} \tag{34}$$

Patients are assigned to groups (35) subject to the groups' capacity constraints (36):

$$\sum_{q \in Q_m} v_{pqt}^{(3)} = x_{pmt}^{(3)} \quad p \in P, m \in M_p^{(2,3)}, k_m > 1, t \in T, x_{pmt}^{(2)} = 1 \tag{35}$$

$$\sum_{\substack{p \in P_m^{(2,3)} \\ x_{pmt}^{(2)}=1}} v_{pqt}^{(3)} \leq k_m \quad m \in M, k_m > 1, q \in Q_m, t \in T, z_{qt}^{(2,3)} = 1 \tag{36}$$

Exogenously scheduled treatments (37), group loyalty (38) and resource loyalty (39) are enforced:

$$x_{pmt}^{(3)} = 1 \quad p \in P, t \in T, m \in GM_{pt}^{(2,3)}, x_{pmt}^{(2)} = 1 \tag{37}$$

$$v_{pqt}^{(3)} = 1 \quad p \in P, t \in T, m \in GM_{pt}^{(2,3)}, q \in lQ_{pm} \tag{38}$$

$$y_{pmft}^{(3)} = x_{pmt}^{(3)} \quad p \in P, f \in F, m \in M_p^{(2,3)} \cap lM_{pf}, k_m = 1, t \in T, x_{pmt}^{(2)} = 1 \tag{39}$$

Note that it is possible that a treatment is assigned to a particular day in the solution of the Stage 1 model and (tentatively) scheduled to start at a particular time slot t , but cannot finally be scheduled in the solution of the Stage 3 model. Such a situation can occur if the availability of several individual resources of the same category is fragmented over the time slots t . In such a case the aggregate capacity in Stages 1 and 2 might (wrongly) seem to be sufficient to schedule a treatment to start at a time slot t which is only seen to be impossible as the Stage 3 model is solved.

4 Numerical experiments

4.1 Outline of the numerical experiments

We performed a number of numerical experiments with the models presented in section 3 to answer the following four questions: First, we wanted to explore the computational effort to solve both the monolithic model and the hierarchical model system. The question is how the computational effort is affected by the size of the problem instance as determined by the number of patients, resources, prescriptions and the time horizon. Second,

we were interested in the loss of solution quality of the hierarchical model system compared to the monolithic model that is caused by the hierarchical decomposition and the aggregation of time and resource capacities. Of course, this comparison is only possible for problem instances that can be solved using either model. Third, we wanted to know how the answers to the first two questions depend on the utilization of the resources. Fourth, we studied the impact of exogenous settings on the solution times and expected that c.p. solution times decrease as the fraction of exogenous settings increase. For our experiments we used a 3GHz Intel Core 2 Quad Pentium PC with 8 GB RAM, under a 64-bit version of Windows 7. The model was implemented in GAMS 23.3.3. CPLEX 12.1.0 was used to solve the models.

4.2 Description of the test cases

The average number of beds in rehabilitation hospitals in Germany in 2008 was 138, with an average bed utilization of 81%. The average length of stay of inpatients was 25 days, see Statistisches Bundesamt Deutschland (2010). However, there is substantial variance in the number of beds per hospital. Furthermore, many rehab hospitals treat a substantial number of outpatients who only come to the hospital for treatments as specified by their individual schedule. This increases the variance of the number of patients to be scheduled. We used data on activities and prescriptions collected from different rehab hospitals and made anonymous by a software provider serving those hospitals.

In our numerical study, we worked with two different test instances modeling virtual rehab hospitals. Aggregate information about the features of these test instances is provided in Table 1. A subset of the treatments, in particular those requiring psychotherapists, were assumed to require resource loyalty with respect to the therapists. For those treatments patients were assigned to individual therapists. The number of active prescriptions is the number of events to be scheduled, i.e., the sum of the prescribed replications over all patient-treatment combinations.

Instance I1 reflects the situation of a rather small and possibly very specialized rehab hospital with a limited number of patients and resources used in the rehab processes. Scheduling the treatments in this rehab hospital is a manageable task for which the monolithic model (still) can be solved. The second problem instance I2 models the situation in a substantially larger institution with more patients, therapists etc.

In order to examine the impact of the resource utilization, we furthermore defined in Table 2 three different utilization scenarios based on the following two quantities: The gross utilization $util_{res}^{grs}$ per resource category \hat{f} is the workload due to prescribed individual treatments plus the workload due to exogenously scheduled group activities divided by the total available capacity of that resource category. The net utilization $util_{res}^{net}$ per resource category is defined similarly, but based on finally scheduled treatments. If all prescribed treatments can be scheduled, gross and net utilization are equal.

Table 1: Key data of the test instances

Indicator \ Test instance	I1 (Instance 1)	I2 (Instance 2)
Number of days $ D $	30	30
Time slots per day $ T $	120	120
Number of resource categories $ \hat{F} $	18	26
Number of resources $ F $	18	92
Number of patients $ P $	40	429
Number of activities $ M $	25	182
Resource-loyal activities	3	46
Number of groups $ Q $	17	224
Number of patient-loyal groups	1	2
Number of precedence relationships for activities	1	6
Number of active prescriptions (base case)	2,458	36,301
Resource-loyal prescriptions (base case)	40	1,349

In addition to the utilization of resources at the level of resource categories, we also considered the gross utilization $util_{grp}^{grs}$ of the exogenously scheduled groups $q \in Q_m$ for collective activities $m, \forall k_m > 1$. The ratio of the duration of the prescribed group activities to the total capacity of the groups dedicated to the specific activities yields the gross group utilization of the groups for activity m . The net group utilization $util_{grp}^{net}$ is determined by computing this ratio for actually scheduled activities of patients participating in these groups.

Table 2: Utilization scenarios

Scenario	L (Low)	M (Medium, base case)	H (High)
Instance I1			
Active prescriptions	1,570	2,458	3,394
Gross resource utilization $util_{res}^{grs}$ [%]	[29%-73%]	[45%-88%]	[60%-120%]
Weighted resource utilization [%]	57%	76%	100%
Gross group utilization $util_{grp}^{grs}$ [%]	[45%-53%]	[68%-78%]	[91%-111%]
Weighted group utilization [%]	50%	78%	107%
Instance I2			
Active prescriptions	23,471	36,301	49,824
Gross resource utilization $util_{res}^{grs}$ [%]	[46%-82%]	[70%-86%]	[77%-118%]
Weighted resource utilization [%]	66%	81%	99%
Gross group utilization $util_{grp}^{grs}$ [%]	[44%-80%]	[67%-86%]	[80%-119%]
Weighted group utilization [%]	52%	80%	111%

Table 2 presents the range of the gross utilizations over all resource categories and exogenously scheduled groups in each scenario for the entire planning horizon. In the utilization scenario H (High) we studied the case in which physicians prescribed more treatments than the rehab hospital actually can carry out. To construct the different utilization scenarios, we initially allocated the number of resource units in each category necessary to reach the medium weighted utilization around 76% for the 2,458 active prescriptions of the I1 base case and around 81% for the 36,301 active prescriptions of the

I2 base case. Afterwards we scaled the number of prescriptions up by a factor $\frac{110}{80}$ and down by a factor $\frac{50}{80}$ and rounded to integer values to reach the “High” and “Low” utilization as given in Table 2 for both the groups and the resource categories. In Table 2 we present the resulting number of active prescriptions as well as the prescription-weighted average of gross resource and group utilization for these active prescriptions for each combination of instance and utilization scenario.

4.3 Numerical results

Tables 3 and 4 present the results of our numerical study in an aggregate form. In Table 3 we first study Instances I1 and I2 without considering exogenously scheduled single treatments. The upper part gives the results for the solution of the monolithic model. The key performance measure is the number of scheduled treatments (ST). We furthermore report the number of rows and columns of the matrix *after* the matrix-reduction preprocessing step performed by CPLEX prior to the optimization, the size (in MB) of the matrix generated by GAMS and transferred to CPLEX, the CPU time CPLEX required to solve the model, the optimality gap when the optimization terminated, and the average utilization of resource categories as well as of groups. The information provided for the Stage 1 model is similar. The models for Stages 2 and 3, however, treat individual days with different workloads as determined at Stage 1. We therefore provide average information for the different quantities over all 30 models of the planning horizon (30 days). In the bottom part of Table 3, we finally report data on the solution of the hierarchical model system (HMS) for the complete planning horizon that is created by combining the results of the Stage 3 models for the 30 different days. We worked with a CPU time limit of one hour for the monolithic model and of 100, 400 and 100 seconds for Stages 1, 2 and 3 of the hierarchical system. In either case we terminated the solution when an integrality gap of 1% was reached.

The results in Table 3 show that the monolithic model can solve the small I1 instances, but not the larger I2 instances. (The required storage for the matrices of the monolithic model in the I2 cases exceeded the 8 GB of RAM available on our computer.) The hierarchical system, however, can solve both instances. The Stage 1 and Stage 3 models tend to be solved quickly while the Stage 2 model of the hierarchical system appears to be relatively hard to solve. For the larger I2 instances we observed several cases where the Stage 2 model was not solved to an integrality gap of 1% within the given time limit of 400 seconds. Note that our aggregation approach limits the size of the models at Stages 1 to 3 substantially—compared to the model sizes of the monolithic model.

We made a very interesting observation by comparing the number of rows and columns of the matrix prior to and after the matrix-reduction preprocessing step performed by CPLEX. For the monolithic model, the preprocessing step eliminated more than 90% of both the rows and columns. For the Stage 1 model, typically 40%-50% of the rows and

Table 3: Model statistics (No single treatments scheduled exogenously)

Instance & Scenario	I1-L	I1-M	I1-H	I2-L	I2-M	I2-H
Active prescriptions	1,570	2,458	3,394	23,471	36,301	49,824
Monolithic model						
Rows	100,841	101,407	101,472			
Columns	302,087	301,425	300,020			
Size [MB]	2,026	1,933	1,878			
CPU [sec.]	219	3,625 (limit)	3,624 (limit)			
Gap [%]	0.0	4.6	7.6			
ST	1,467	2,108	2,454			
$util_{res}^{net}$ [%]	54	64	68			
$util_{grp}^{net}$ [%]	45	70	85			
Stage 1 model						
Rows	7,033	7,355	7,188	109,597	120,184	119,967
Columns	9,212	9,199	9,147	184,228	184,215	184,045
Size [MB]	15	14	13	219	196	181
CPU [sec.]	0.4	0.4	0.4	15.7	22.4	52.3
Gap [%]	0.1	0.2	0.4	0.0	0.1	0.3
ST	1,496	2,346	2,936	23,130	35,629	44,299
Stage 2 models						
Avr. Rows	499	746	999	10,885	17,895	22,936
Avr. Columns	1,461	2,390	3,026	33,136	51,095	63,842
Avr. Size [MB]	15	19	20	187	233	259
Avr. CPU [sec.]	0.1	0.2	0.3	98.8	205.5	402.0
Avr. Gap [%]	0.0	0.0	0.0	0.9	1.8	3.7
Avr. ST	45	69	83	747	1,137	1,375
Stage 3 models						
Avr. Rows	all elim.	all elim.	all elim.	1,514	2,539	2,886
Avr. Columns	all elim.	all elim.	all elim.	2,318	3,672	4,391
Avr. Size [MB]	10	11	11	115	122	121
Avr. CPU [sec.]	0.0	0.0	0.0	0.2	0.1	0.2
Avr. Gap [%]	0.0	0.0	0.0	0.0	0.0	0.0
Avr. ST	45	69	83	704	1,069	1,294
Final solution HMS						
ST	1,335	2,077	2,502	21,112	32,056	38,805
$util_{res}^{net}$ [%]	49	65	72	56	66	70
$util_{grp}^{net}$ [%]	44	69	85	50	77	94

Table 4: Model statistics (30% patients with exogenously scheduled single treatments)

Instance & Scenario	I1-L	I1-M	I1-H	I2-L	I2-M	I2-H
Active prescriptions	1,570	2,458	3,394	23,471	36,301	49,824
Monolithic model						
Rows	65,815	60,246	50,340			
Columns	184,393	168,353	134,193			
Size [MB]	2,026	1,933	1,878			
CPU [sec.]	119.3	3,625 (limit)	951.4			
Gap [%]	0.1	1.4	0.5			
ST	1,465	2,163	2,550			
$util_{res}^{net}$ [%]	54	66	70			
$util_{grp}^{net}$ [%]	45	71	87			
Stage 1 model						
Rows	6,125	6,594	6,562	96,449	105,772	107,029
Columns	8,177	8,270	8,284	165,070	165,077	167,021
Size [MB]	15	14	13	220	196	182
CPU [sec.]	0.4	0.4	0.4	9.0	16.2	33.6
Gap [%]	0.0	0.2	0.6	0.0	0.0	0.2
ST	1,497	2,347	2,930	23,126	35,643	44,302
Stage 2 models						
Avr. Rows	255	344	371	7,888	12,697	15,669
Avr. Columns	656	914	909	24,227	35,972	43,168
Avr. Size [MB]	15	18	21	188	235	261
Avr. CPU [sec.]	0.1	0.1	0.1	60.0	151.8	386.5
Avr. Gap [%]	0.0	0.0	0.0	0.7	0.7	1.7
Avr. ST	44	68	82	748	1,144	1,391
Stage 3 models						
Avr. Rows	all elim.	all elim.	all elim.	1,524	2,547	2,911
Avr. Columns	all elim.	all elim.	all elim.	2,243	3,580	4,286
Avr. Size [MB]	10	11	11	115	123	123
Avr. CPU [sec.]	0.0	0.0	0.0	0.1	0.2	0.2
Avr. Gap [%]	0.0	0.0	0.0	0.0	0.0	0.0
Avr. ST	44	68	82	706	1,079	1,314
Final solution HMS						
ST	1,321	2,049	2,459	21,175	32,356	39,428
$util_{res}^{net}$ [%]	49	64	71	56	67	72
$util_{grp}^{net}$ [%]	43	68	82	50	77	95

5%-20% of the columns were eliminated. In case of the Stage 2 model, in most cases 50% to 90% of both the rows and columns were eliminated. At Stage 3 the preprocessor eliminated *all* rows and columns for Instance I1. The reason is that in this instance there was always just one resource per resource type so that there was eventually nothing left to decide with respect to the concrete resource used for any scheduled treatment. In case of the Instance I2, around 70% of the rows and of the columns were eliminated.

The computation times of the hierarchical model system are substantially lower than those of the monolithic model. In many real-world settings, the computation times of the monolithic model would be prohibitively long. As expected, the monolithic model provides in principle better solutions, i.e., more scheduled treatments. The loss of 5% to 10% of the treatments that are not scheduled due to the hierarchical decomposition can be noticeable. Note that in the Stage 1 model more treatments are scheduled than in the solution of the monolithic model as well as the final solution of the hierarchical model system because of the aggregate time and resource constraints used in the Stage 1 model.

If the resource utilization increases, the computation times of the monolithic model and the Stage 2 model tend to increase as well. It is also interesting that the resource utilization in the I1-H and I2-H cases is substantially below 100%, despite of the up-scaling of the active prescriptions. Here the time and stress constraints of the patients apparently permitted to schedule many of the additional treatments (compared to the base case).

We finally asked how the performance of both systems is affected if a substantial number of treatments is exogenously scheduled with respect to time and resource assignment. In order to answer this question we took the results from the first run that led to the results in Table 3 and randomly selected 30% of the patients. For these 30% of the patients, we exogenously fixed the schedule from the first run of the system and started the optimization again. This way we could easily ensure that all the exogenous settings were capacity-feasible and reduce the solution space as 30% of the solution was already given. The results are reported in Table 4. They show that in all cases a very similar number of total treatments was scheduled when the CPU time limit or the integrality gap limit was reached. On average the computational effort decreased (as expected) by exogenously scheduling a large fraction of the treatments. However, we also observed some cases in which the computation times increased slightly. Apparently the search tree of the branch&bound process to solve the models can exhibit a disadvantageous structure leading to longer computation times if some of the decision variables are fixed beforehand.

It hence appears that it is possible to automatically schedule treatments in a rehab hospital using our hierarchical model system. It might be a valuable contribution for the development of software and for the management of rehab hospitals.

5 Conclusions and outlook

In this paper we presented the central elements of a model-based decision support system for patient and resource scheduling in rehabilitation hospitals. The scheduling problems in rehabilitation hospitals appear to have features that make the use of a formal system both possible and worthwhile. The size of the scheduling instances appears to demand a hierarchical scheduling approach as developed in this paper. Our hierarchical approach has the attractive feature that the near future within the planning horizon can be treated in detail solving the Stage 2 and 3 models, while the more distant future can be treated at an aggregate level only in the Stage 1 model. Furthermore, the Stage 2 and 3 models can be solved in isolation for different days. This gives the possibility, for example, to re-schedule the plan for the next day (in isolation) if a patient or a therapist becomes ill and his appointments therefore have to be canceled or re-assigned.

The numerical results show that the approach works well in principle for virtual rehab hospitals of a substantial size. It would certainly be both necessary and possible for a real-world application to add constraints which reflect specific requirements of the particular rehab hospital. Our scheduling approach—as based entirely on algebraic models and the use of commercial solvers—appears to be most flexible and hence well-suited to adapt to these needs.

Using such a system would substantially change the processes in—and the management of—a rehab hospital. The system would be a useful tool in the discussions between physicians and schedulers about treatments to prescribe and resources to allocate in order to serve the patients well and to operate efficiently at the same time. Schedulers would focus less on scheduling individual treatments and more on decisions about the resource usage on the global level, e.g., by deciding about overtime, creation of additional groups etc. We plan to address these questions in our further work.

References

- Beliën, J. and E. Demeulemeester (2007). Building cyclic master surgery schedules with leveled resulting bed occupancy. *European Journal of Operational Research* 176(2), 1185–1204.
- Beliën, J., E. Demeulemeester, and B. Cardoen (2009). A decision support system for cyclic master surgery scheduling with multiple objectives. *Journal of Scheduling* 12(2), 147–161.
- Brandeau, M. L., W. P. Pierskalla, and F. Sainfort (Hrsg.) (2005). *Operations Research and Health Care: A Handbook of Methods and Applications*, Band 70 of *International Series in Operations Research & Management Science*. Boston, MA: Springer Science + Business Media Inc.
- Brunner, J. O. (2010). Flexible Shift Planning in the Service Industry: The Case of Physicians in Hospitals.

- Brunner, J. O., J. F. Bard, and R. Kolisch (2009). Flexible shift scheduling of physicians. *Health Care Manag Sci* 12(3), 285–305.
- Cardoen, B., E. Demeulemeester, and J. Beliën (2010). Operating room planning and scheduling: A literature review. *European Journal of Operational Research* 201(3), 921–932.
- Cheang, B., H. Li, A. Lim, and B. Rodrigues (2003). Nurse rostering problems—a bibliographic survey. *European Journal of Operational Research* 151(3), 447–460.
- Chien, C. F., Y. C. Huang, and C. H. Hu (2009). A hybrid approach of data mining and genetic algorithms for rehabilitation scheduling. *Int. J. Manuf. Technol. Manage.* 16(1-2), 76–100.
- Chien, C.-F., F.-P. Tseng, and C.-H. Chen (2008). An evolutionary approach to rehabilitation patient scheduling: A case study. *European Journal of Operational Research* 189(3), 1234–1253.
- Guerriero, F. and Rosita Guido (2011). Operational research in the management of the operating theatre: a survey. *Health Care Management Science* 14(1), 89–114.
- Hall, R. W. (Hrsg.) (2006). *Patient Flow: Reducing Delay in Healthcare Delivery*, Band 91 of *International Series in Operations Research & Management Science*. Berlin, Heidelberg: Springer.
- Kellogg, D. L. and S. Walczak (2007). Nurse Scheduling: From Academia to Implementation or Not? *Interfaces* 37(4), 355–369.
- Kolisch, R. and S. Sickinger (2008). Providing radiology health care services to stochastic demand of different customer classes. *OR Spectrum* 30(2), 375–395.
- Ogulata, S. N., M. Koyuncu, and E. Karakas (2008). Personnel and Patient Scheduling in the High Demanded Hospital Services: A Case Study in the Physiotherapy Service. *Journal of Medical Systems* 32(3), 221–228.
- Podgorelec, V. and P. Kokol (1997). Genetic Algorithm Based System for Patient Scheduling in Highly Constrained Situations. *Journal of Medical Systems* 21(6), 417–427.
- Sickinger, S. (2008). *Optimierung der Terminvergabe und Patientenauswahl in Radiologischen Instituten von Krankenhäusern*. Aachen: Shaker Verlag.
- Statistisches Bundesamt Deutschland (2010). Vorsorge- oder Rehabilitationseinrichtungen. Einrichtungen, Betten und Patientenbewegung. www.destatis.de.
- Vanberkel, P. T., R. J. Boucherie, E. W. Hans, J. L. Hurink, and N. Litvak (2010). A Survey of Health Care Models that Encompass Multiple Departments. *International Journal of Health Management and Information* 1, 37–69.